

Aspects mécanistiques et énergétiques des interactions entre l'ADN et une molécule intercalante

THESE de DOCTORAT

soutenue publiquement le 13 Juillet 2012

en vue de l'obtention du

Diplôme de doctorat de l'Université Claude Bernard de Lyon
Spécialités Aspects Moléculaires et Cellulaires de la Biologie

Ecole Doctorale - EDISS

(arrêté du 7 août 2006)

par

Matthieu WILHELM

Directeur de Thèse : Richard LAVERY

Rapporteurs : Chantal PREVOST
Yves-Henri SANEJOUAND
Damien LAAGE

Examineur Gilbert DELEAGE

Laboratoire de Bioinformatique : Structures et Interactions (BISI)

Bases Moléculaires et Structurales des Systèmes Infectieux (BMSSI)

IBCP UMR 5086, 7 passage du Vercors, 69007 Lyon, France

Remerciements

Je tiens tout d'abord à adresser mes plus vifs remerciements au Docteur Richard Lavery mon directeur de thèse. Son enthousiasme sans faille, sa passion et sa grande compétence scientifique ont été un réel moteur tout au long de ces travaux de thèse.

Je remercie Madame Chantal Prévost et Messieurs Yves-Henri Sanejouand et Damien Laage d'avoir aimablement accepté d'être les rapporteurs de cette thèse. Merci également à Monsieur Gilbert Deléage qui a bien voulu examiner mon travail.

Merci à Benjamin Bouvier qui m'a apporté beaucoup au jour le jour notamment au niveau scientifique et technique, mais également au cours de ces longs échanges de point de vu sur les sujets les plus divers et variés. Vive le « hop là land !! ».

Une pensée pour Krystyna dont les discussions durant nos petites pauses, ont toujours été des plus enrichissantes ; et également pour Juliette dont les conseils se sont toujours avérés des plus avisés durant ces trois ans.

Un énorme Merci à l'Italie, qui au-delà de posséder des spécialités culinaires sans comparaison, nous a envoyé deux de ses meilleurs éléments. – Nicco, avec qui on a pu partager nos expériences réciproques durant ces trois longues années ! Tu vois, finalement on y est arrivé ! – Marco, je tiens à te remercier non seulement pour tous les bons moments (et il y en eu beaucoup), mais aussi et surtout pour les rudiments d'italien que tu as su enseigner à tout le labo (je vais peut-être éviter de les répéter ici...). Café ??

Egalement un grand merci à monsieur Alexis. Je pense que tu sais déjà tout le bien que je pense de toi, comme à peu près la moitié de l'institut d'ailleurs... Pour l'autre moitié, ce ne sont que des jaloux de ta nouvelle coupe !

Une énorme pensée pour ma ptite Tetette (j'suis sûr que tu l'avais oublié celui-ci), qui a su prendre son rôle de coloc de thèse très au sérieux et surtout à su me supporter dans les pires moments de pétage de plomb. Ça va me manquer de ne plus t'avoir à côté de moi tous les jours.

Merci à ma Wiki-Fanny, dont le départ du labo a laissé un grand vide. T'as eu beau nous répéter que t'étais là aussi pour travailler, je t'assure que depuis ton départ, le labo est devenu « très » calme... On se refait un été comme celui de 2009 quand tu veux !

Une mention spéciale pour ma marseillaise préférée, à la spontanéité et à l'énergie plus que débordante. Merci d'avoir partagé ces moments intenses de fin de thèse avec moi. Encore un petit effort, tu y es presque toi aussi ! Pense à la carte du monde que tu as à finir. Et surtout, ne change rien !

Une petite pensée également à toutes les personnes que j'ai pu apprécier au sein de l'institut pendant ces trois années et demi ; aussi bien les gens d'en face au troisième que les asociaux du deuxième et tous les autres que j'oublie. Une mention spéciale pour nos gestionnaires et plus particulièrement à Kitou et Dorothée qui savent nous rendre la vie beaucoup plus facile.

Un grand merci à toutes les personnes qui ont su rendre ma vie lyonnaise si agréable et surtout au Gamin et à Biggi (que j'aimais à décrire comme « On dirait un peu Mario qui a pris une étoile »). Merci à vous deux pour tous ces moments de vie commune, toujours facile, et au final trop court. Quand vous voulez vous passez à la maison, le balcon est prêt et la bière est au frais.

Un merci à Lillie, non seulement de m'avoir permis de rencontrer mon Rayon de Soleil, mais surtout pour tous ces bons moments à discuter et ces fous rires. Je te souhaite que du bonheur. Et n'oublie pas, on a toujours un bad sur le feu.

Une mention spéciale à Nespresso qui a rythmé mon quotidien tout au long de ces travaux, et sans qui mes journées auraient été bien plus longues à démarrer.

Une pensée également pour mes amis de « pour la vie » Jo, Loïc et Marc, qui malgré la distance ont toujours su rester proche de moi tout au long de mes études.

Merci à toi Vincent. Depuis notre lointaine ville de Metz, les hasards de la vie ont toujours permis qu'on se trouve au même endroit quasiment en même temps. Je pense qu'au fil du temps, que tu le veuilles ou non, tu t'es avéré un peu être le frère qu'il me manquait. Ta chambre est prête, donc quand tu veux tu passes à la maison.

Je tiens également à adresser mes remerciements à ma future « jolie » famille qui m'a très vite accepté et intégré à la sympathique petite tribu Calamel. T'as vu Lu, finally we did it !!

Des remerciements tous particuliers pour ma famille. Tout d'abord à mes sœurs, qui même si la distance nous a séparé, ont toujours gardé un regard bienveillant sur leur petit frère. Aussi et surtout pour mes parents, qui malgré les quelques échecs qui ont parfois jalonné mon parcours ainsi que parfois le manque de compréhension vis à vis de ce que je faisais vraiment, m'ont toujours apporté leur soutien inconditionnel. Je pense que je ne vous remercierai jamais assez.

Finalement, je ne peux finir ces remerciements sans te remercier toi, Astrid ! Je t'ai promis six mots mais tu mériterais six chapitres à toi toute seule. Merci pour ton soutien de tous les jours (et je sais que ça a pas du être toujours facile à la fin), merci d'avoir été là dans les bons et les mauvais moments, merci d'avoir partagé tout ça avec moi, merci pour ta compréhension et ta patience à toutes épreuves et surtout merci pour ton amour ! Cette thèse je te la dois aussi à toi ! Promis, à partir de maintenant on profite !

Résumé

L'ADN est au cœur de nombreux processus biologiques. De ce fait, il est la cible de nombreuses molécules pharmacologiques employées notamment dans les thérapies anticancer. Parmi celles-ci, la daunomycine va interagir avec la double hélice d'ADN en s'intercalant entre deux paires de bases, bloquant ainsi la réplication. Malgré l'efficacité reconnue des molécules intercalantes, et de nombreuses études à ce sujet, le mécanisme du processus d'intercalation n'est pas clairement déterminé à l'heure actuelle.

Basés sur l'utilisation de la dynamique moléculaire avec une représentation tout atome des systèmes en conditions de solvant explicite, ces travaux ont tout d'abord visé à caractériser l'influence de la flexibilité conformationnelle de la daunomycine vis à vis de l'ADN, ainsi que son importance dans les interactions entre ces deux molécules.

Dans un deuxième temps nous nous sommes focalisés sur le chemin réactionnel menant la daunomycine à son site d'intercalation par umbrella sampling. Cette étude, en plus de fournir des énergies libres en accord avec les données expérimentales, a permis de pointer du doigt un mécanisme d'intercalation impliquant une étape préliminaire de liaison de la daunomycine au petit sillon de l'ADN, suivie d'une étape intermédiaire de réorientation du ligand.

Finalement, la réalisation de simulations de glissement de la daunomycine le long du petit sillon de l'ADN nous a permis de nous intéresser au mécanisme permettant à ce ligand de localiser son site d'intercalation le long de l'ADN.

Table des matières

Introduction Générale	15
Chapitre 1 L'ADN	17
1.1 Structure et organisation de l'ADN	19
1.1.1 Les nucléotides	20
1.1.2 Structure de la double hélice d'ADN	22
1.2 Analyse conformationnelle	22
1.2.1 Flexibilité de la double hélice	23
1.2.2 Paramètres hélicoïdaux	26
1.2.3 Les sillons de l'ADN	27
1.2.4 Conformations de l'ADN	29
1.3 Les effets de séquences	33
1.4 Propriétés dynamiques et mécaniques de l'ADN	36
1.4.1 De l'eau et des ions	36
1.4.2 Polymorphisme conformationnel et influence du milieu	41
Chapitre 2 Molécules intercalantes	45
2.1 Modes d'interactions avec l'ADN	47
2.1.1 Modes d'interactions	48
2.1.1.1 Liaisons aux sillons de l'ADN	48
2.1.1.2 Liaisons au squelette sucre phosphate	50
2.1.1.3 Intercalation entre deux paires de bases	51
2.2 Intercalation et molécules intercalantes	51

2.2.1	Intercalation « classique »	52
2.2.1.1	Principe	52
2.2.1.2	Déformation de l'ADN	52
2.2.1.3	Le principe d'exclusion du plus proche voisin	53
2.2.2	Autres types d'intercalants et d'intercalations	54
2.2.2.1	Bis-intercalation	54
2.2.2.2	Threading intercalation	55
2.2.3	Intérêts biochimiques et médicaux	56
2.2.3.1	Intercalants et thérapie anticancer	56
2.2.3.2	Marquage	57
2.2.3.3	Effets biologiques des anthracyclines – inhibition des topoisomérases	58
2.3	La daunomycine	59
2.3.1	Structure libre et intercalée	60
2.3.2	Flexibilité du cycle A	62
2.4	Aspects thermodynamiques et cinétiques	64
 Chapitre 3 Méthodes		 69
3.1	D'une vision rigide à une vision dynamique : introduction à la dynamique moléculaire	71
3.2	Champs de forces	73
3.2.1	Expression du champ de forces	73
3.2.2	Interactions intramoléculaires – « liées »	73
3.2.3	Interactions intermoléculaires – « non liées »	75
3.2.3.1	Interactions de Van der Waals	75
3.2.3.2	Interactions électrostatiques	75
3.2.4	De la formulation à l'utilisation pour l'ADN	76
3.3	Traitement du solvant et des ions	77
3.3.1	Quelle représentation de l'environnement ?	77
3.3.1.1	Représentation implicite	78
3.3.1.2	Représentation explicite	78
3.3.2	Modèles de représentation de l'eau	79

3.3.3	Les ions	81
3.4	Simulation de dynamique moléculaire	82
3.4.1	Minimisation de l'énergie	83
3.4.2	Dynamique moléculaire	85
3.4.2.1	Dynamique Newtonienne	85
3.4.2.2	Dynamique de Langevin	86
3.4.2.3	Méthodes d'intégrations	87
3.4.2.4	Algorithme SHAKE	88
3.4.3	Ensembles de simulations	89
3.4.4	Conditions initiales de simulations	90
3.4.5	Dynamiques moléculaires à température et pression...	91
3.4.5.1	Contrôle de la température	91
3.4.5.2	Contrôle de la pression	93
3.4.6	Conditions aux limites	95
3.4.7	Evaluation des interactions non liées	96
3.4.7.1	Troncature sphérique	96
3.4.7.2	Traitement des interactions à longues distances	99
3.5	Protocoles généraux de simulations	104
3.6	Energie libre	106
3.6.1	Calculs des différences d'énergie libre	107
3.6.1.1	Perturbations thermodynamiques	108
3.6.1.2	Intégrations thermodynamiques	109
3.6.1.3	Méthode de « croissance lente » - (« slow growth » method)	109
3.6.2	Potentiels de force moyenne - PMF	110
3.6.3	Umbrella sampling	111
3.6.3.1	Description	111
3.6.3.2	Weighted Histogram Analysis Method – WHAM	113
3.6.3.3	Analyses statistiques	114
3.6.3.4	Métadynamique	115
3.7	Outils d'analyses	118
3.7.1	Analyses structurales	118
3.7.1.1	Logiciels d'analyse	118
3.7.1.2	RMSD	118

3.7.1.3	RMSD-2D	119
3.7.1.4	Génération de structures moyennes	119
3.7.2	Curves+	120
3.7.2.1	Principes	120
3.7.2.2	Extension de Curves+ à la position du ligand	123
Chapitre 4 Etude de la Flexibilité Interne de la Daunomycine		125
4.1	Construction des systèmes	129
4.2	Paramètres hélicoïdaux de l'ADN	132
4.3	Le cycle A de la daunomycine	134
4.3.1	Description du cycle A en fonction de son environnement	134
4.4	Etude des transitions conformationnelles du cycle A	137
4.4.1	DNM libre en milieu aqueux	138
4.4.2	DNM intercalée	140
4.4.2.1	Etudes des chemins de transition du cycle A en...	140
4.4.2.2	Etude des chemins de transition en conditions de sels...	146
4.4.2.3	Chemins de transitions et effets de la concentration ionique	149
4.4.3	DNM liée au petit sillon de l'ADN	150
4.4.3.1	Transitions du dièdre τ	150
4.4.3.2	Influence de la concentration ionique	152
4.4.3.3	Etudes des paramètres hélicoïdaux de l'ADN le long des...	155
4.5	Discussion et conclusions	158
Chapitre 5 Etude du chemin de désintercalation		161
5.1	Introduction	163
5.2	Description de la contrainte de séparation	165
5.3	Chemin de passage de la DNM de son site d'intercalation au petit...	166
5.3.1	Séparation à partir de la position IC	166
5.3.1.1	Profil d'énergie libre	168

5.3.1.2	Description du chemin de séparation	171
5.3.2	Séparation à partir de la position MG1	173
5.3.2.1	Profil d'énergie libre du chemin de séparation MG1	173
5.3.2.2	Description du chemin de séparation	175
5.3.3	Passage de la DNM de son site d'intercalation au petit sillon de l'ADN	177
5.4	Caractérisation du passage de la DNM de son site d'intercalation au petit sillon de l'ADN par métadynamique	181
5.4.1	Surface d'énergie libre	181
5.4.2	Chemin de séparation	184
5.5	Comparaison des deux chemins	184
5.6	Discussion et conclusion	191

Chapitre 6 Glissement de la DNM le long du petit sillon de l'ADN 193

6.1	Un point sur les stratégies de localisation et le glissement	195
6.2	Stratégie et contrainte de glissement	198
6.3	Validation du modèle	201
6.3.1	Conditions de simulations et étapes préliminaires	201
6.3.2	Séquence polyG régulière	202
6.3.3	Séquence polyG avec <i>rise</i> alternée	203
6.4	Simulation du glissement de la DNM dans l'eau	205
6.4.1	Description du système	205
6.4.2	Chemins initiaux	207
6.4.2.1	Profil d'énergie libre	209
6.4.2.2	Cohérence des échantillonnages	210
6.4.3	Chemin avec la conformation interne de la DNM fixée	213
6.4.4	Chemin final	217
6.4.4.1	Profil d'énergie libre	217
6.4.4.2	Cohérence des échantillonnages	218
6.4.4.3	Analyses conformationnelle	219

Table des matières

6.4.5	Du glissement à l'intercalation	223
6.5	Discussion et conclusions	225
Conclusions Générale		227
Annexe A Evolution des paramètres hélicoïdaux		231
Annexe B Article publié		237
Annexe C Recouvrement des échantillonnages		249
Bibliographie		251

Introduction Générale

A l'heure de la profusion des données biologique, acquise à l'aide de méthodes de plus en plus performantes, on ne peut que s'émerveiller devant l'acharnement du vivant à cacher tous ses secrets. La compréhension des mécanismes guidant son fonctionnement s'avère souvent aussi fastidieuse que passionnante. Dans un tel contexte, l'étude des processus d'interactions entre les macromolécules ressemble à s'y méprendre, à une chasse au trésor menée avec une carte incomplète. On sait à quoi doit ressembler ce trésor (grâce aux structures RMN et cristallographique), mais la direction et le chemin permettant de le rejoindre manquent cruellement de détails. Muni de la modélisation moléculaire pour seul outil, nous sommes donc partis en quête du mécanisme permettant à une molécule intercalante de l'ADN, de passer du milieu cellulaire, à une position intercalée entre deux paires de bases.

Au sein de ces intercalants, la daunomycine est une petite molécule, très rigide, positivement chargée employée depuis plus de 60 ans dans le cadre de thérapies anticancer. Bien que le processus d'intercalation de cette molécule soit bien défini d'un point de vue thermodynamique, le détail du mécanisme d'intercalation en lui-même demeure incertain. En effet, cette molécule passerait d'un état libre à un état intercalé entre deux paires de bases CpG par l'intermédiaire d'une phase hypothétique de complexation avec le petit sillon de l'ADN ; mais, pour l'heure, le détail de cette intercalation est très discuté tant sur le plan cinétique que mécanistique. Les données structurales concernant ce complexe, bien qu'existant à profusion, ne nous offrent aucune vision dynamique du processus à l'échelle moléculaire. Cette lacune pourra, en partie être surpassée par l'utilisation de la dynamique moléculaire.

Sur ces bases, au cours de ce travail de thèse, nous avons pu considérer les interactions entre la daunomycine et l'ADN sous de nombreux aspects, nous permet-

tant finalement de les appréhender sous un nouveau jour et de proposer une vue d'ensemble du processus d'intercalation.

Ce document est organisé en deux parties. La première partie (comprenant les Chapitres 1 à 3) consiste en une revue de l'état de l'art. Le Chapitre 1 propose une description de la molécule d'ADN sous tous les points de vue. Le Chapitre 2 permet de faire le point sur l'intercalation et la daunomycine, et propose un mécanisme réactionnel en accord avec les données expérimentales. Au cours du Chapitre 3 nous nous sommes concentrés sur la description de tous les outils et algorithmes qui nous ont été nécessaires au cours de ces années de thèses. La seconde partie, composée des Chapitre 4 à 6, est entièrement consacrée à nos travaux traitant des interactions entre la daunomycine et l'ADN. Le Chapitre 4 aborde ces interactions en se focalisant sur les aspects conformationnels. Le Chapitre 5, quant à lui, permet de décrire le processus d'intercalation de la daunomycine entre deux paires de bases de l'ADN d'un point de vue mécanistique, conformationnel et énergétique. Finalement le Chapitre 6 fait état de la première tentative de glissement d'une molécule intercalante le long du petit sillon de l'ADN.

CHAPITRE 1

L'ADN

La première description des acides nucléiques remonte aux travaux du biologiste suisse F. Miescher au XIX^{ème} siècle [1], travaux dans lesquels, suite à la séparation du noyau et du cytoplasme, il isola une substance riche en phosphore et aux propriétés acides qu'il nomma nucléine. S'ensuivirent de nombreux travaux [2-4], ouvrant la voie à la compréhension du rôle fondamental de l'ADN dans le vivant [5-8]. 75 ans après sa première description, et grâce notamment aux travaux par rayon X de Rosalind Franklin et Maurice Wilkins, Francis Crick et James Watson aboutirent à la résolution de la structure de l'ADN [9].

En bientôt 60 ans, le développement des techniques d'analyses structurales (RMN, rayons X, etc) ainsi que l'appréhension d'évènements tels que les interactions, les processus de reconnaissance, les changements conformationnels, la transcription ou encore la réplication, nous ont permis d'asseoir notre compréhension du rôle essentiel de l'ADN au cœur du vivant.

1.1 Structure et organisation de l'ADN

L'ADN est une macromolécule biologique, reconnue pour être le support de l'information génétique. La molécule d'ADN est composée d'un enchainement de nucléotides structurés en brins s'appariant deux à deux, de façon complémentaire en une structure particulière appelée « double hélice ». Chaque nucléotide de l'ADN est composé d'une base (une base azotée) liée à un sucre (qui est un 2'-désoxyribose) par une liaison C1'-N glycosyl, sur lequel est lié au moins un groupement phosphate via une liaison ester.

1.1.1 Les nucléotides

Les bases azotées sont des molécules hétérocycliques aromatiques planes, dérivées structurelles soit des *purines* (R), soit des *pyrimidines* (Y) (Figure 1.1). Les purines utilisées dans l'ADN sont l'*Adénine* (A) et la *Guanine* (G). Quant aux pyrimidines, il s'agit de la *Cytosine* (C) et de la *Thymine* (T). Les purines forment des liaisons avec un sucre à cinq carbones (un pentose) via leur atome N9, tandis que les pyrimidines le font via leur atome N3.

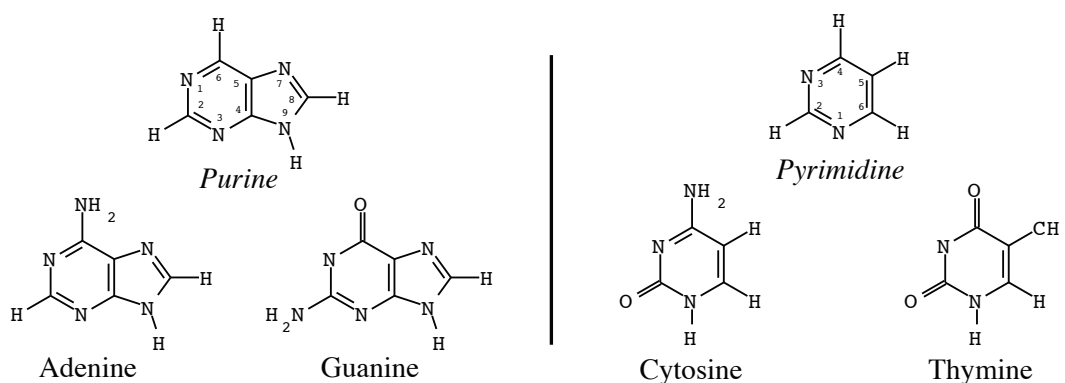


Figure 1.1 - Structure des nucléotides

L'ADN possède un nombre équivalent de résidus adénine et thymine ($A = T$), et un nombre égal de résidus cytosine et guanine ($C = G$). Cette relation connue sous le nom de loi de Chargaff [7], met en évidence la complémentarité des bases puriques et pyrimidiques se reflétant par un appariement des bases via deux ou trois liaisons hydrogènes. Ce sont ces propriétés qui confèrent à l'ADN sa structure particulière en « double hélice », dont la découverte [9] valut à Watson et Crick d'obtenir le prix Nobel en 1962. Dans leur modèle, on retrouve donc deux paires de bases $A \bullet T$ et $C \bullet G$ reliées par l'intermédiaire de liaisons hydrogènes également connues sous le nom de liaison Watson & Crick (Figure 1.2).

Le pentose composant les nucléotides de l'ADN est un 2'-desoxy-D-ribose (ou 2'-desoxyribose). Les groupements phosphates sont liés soit au C3', soit au C5' du 2'-desoxyribose et ainsi forment un 3'-nucléotide ou un 5'-nucléotide respectivement. L'ADN est donc une chaîne de nucléotides, où les phosphates relient les positions 3' et 5' de deux riboses voisins. Cet enchainement permet de définir la structure primaire, ou plus communément, la séquence d'un brin d'ADN ainsi qu'une orientation qui, par convention, est écrite dans le sens 5' → 3'.

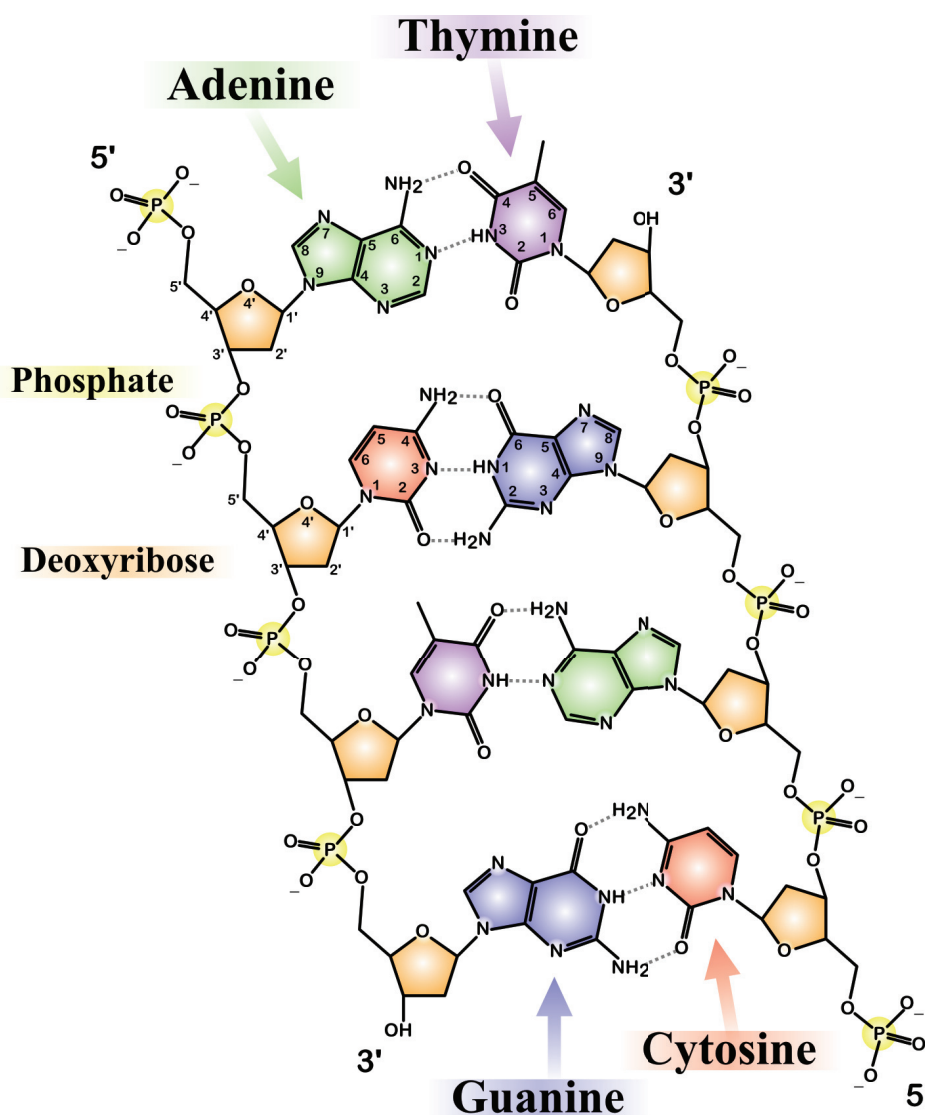


Figure 1.2 – Représentation schématique de la structure de l'ADN, avec enchainement des bases et appariements Watson Crick

1.1.2 Structure de la double hélice d'ADN

Dans la double hélice, les deux brins complémentaires sont orientés en sens contraire, et sont maintenus entre eux grâce aux liaisons hydrogènes entre les paires de bases. Les bases sont empilées perpendiculairement à l'axe de la double hélice, empilement qui va conférer une grande stabilité à la double hélice. Les sucres et les phosphates sont à l'extérieur de cette hélice et forment le squelette. D'un point de vue structurel, chaque tour d'hélice est formé d'environ dix paires de bases (10,5 en solution aqueuse) en 35,7 Å. D'autres paramètres permettant de définir la double hélice (notamment les paramètres hélicoïdaux) seront définis plus loin.

1.2 Analyse conformationnelle

L'ADN ne pourra exprimer ou réguler l'information génétique qu'il contient qu'à travers des interactions avec de nombreux types de molécules différentes allant de la plus simple des molécules ayant un effet pharmacologique à la plus complexe des protéines. Quel que soit le type de molécule interagissant avec l'ADN, cette interaction ne pourra se faire qu'à travers des processus de reconnaissance pouvant impliquer et/ou modifier aussi bien la structure, la charge, la conformation ou encore l'hydratation de la molécule d'ADN. C'est pourquoi l'étude des aspects conformationnels du double brin d'ADN est indispensable à la bonne compréhension des processus nécessaires au fonctionnement de tous les systèmes interagissant avec l'ADN.

Le double brin d'ADN sous son apparente rigidité est en fait une macromolécule flexible, ce qui est due notamment aux nombreuses liaisons simples des brins phosphodiester et à la souplesse des paires de bases.

1.2.1 Flexibilité de la double hélice

Les trois composants chimiques des nucléotides (le groupement phosphate, le désoxyribose et les bases) ont des caractéristiques conformationnelles distinctes qui sont en grande partie responsables de la flexibilité de la double hélice.

Conformations du sucre. Le sucre composant l'ADN est un pentose. Son cycle non plan possède une flexibilité limitée [10,11]. Le désoxyribose peut adopter soit une conformation avec un atome hors du plan dite en *enveloppe* (E), soit une conformation avec deux atomes hors du plan du cycle ou conformation *twist* (T). Le plissement du sucre est défini en fonction de la position de l'atome sortant du plan par rapport au carbone exocyclique C5'. Le plissement sera donc qualifié d'*endo* s'il se trouve du même côté que le C5', et donc *exo* dans le cas contraire. En 1972, les travaux d'Altona & Sundaralingam [10] ont permis de décrire chaque conformation du cycle du sucre par deux paramètres de pseudorotation : l'angle de phase de pseudorotation P, et l'angle maximum de torsion θ_m . Cette description dérive des cinq angles de torsion endocyclique. L'ensemble des conformations adoptées par le cycle furanose peut être placé sur un cercle de pseudorotation (voir Figure 1.3). La phase est considérée égale à 0° lorsque l'angle C1'-C2'-C3'-C4' est maximal, ce qui correspond à la jonction entre les plissements C3'-*endo* et C2'-*exo*. Une phase de 180° correspondant à la jonction entre les plissements C2'-*endo* et C3'-*exo*. Les conformations en enveloppe et en twist sont alternées sur le cercle tous les 18°. Les sucres composant l'ADN adoptent préférentiellement les conformations C3'-*endo* (Nord) et C2'-*endo* (Sud). En solution, les sucres sont généralement en équilibre dynamique entre ces deux conformations, la forme Sud étant la conformation la plus présente en milieux hydratés (correspondant à la forme B de l'ADN).

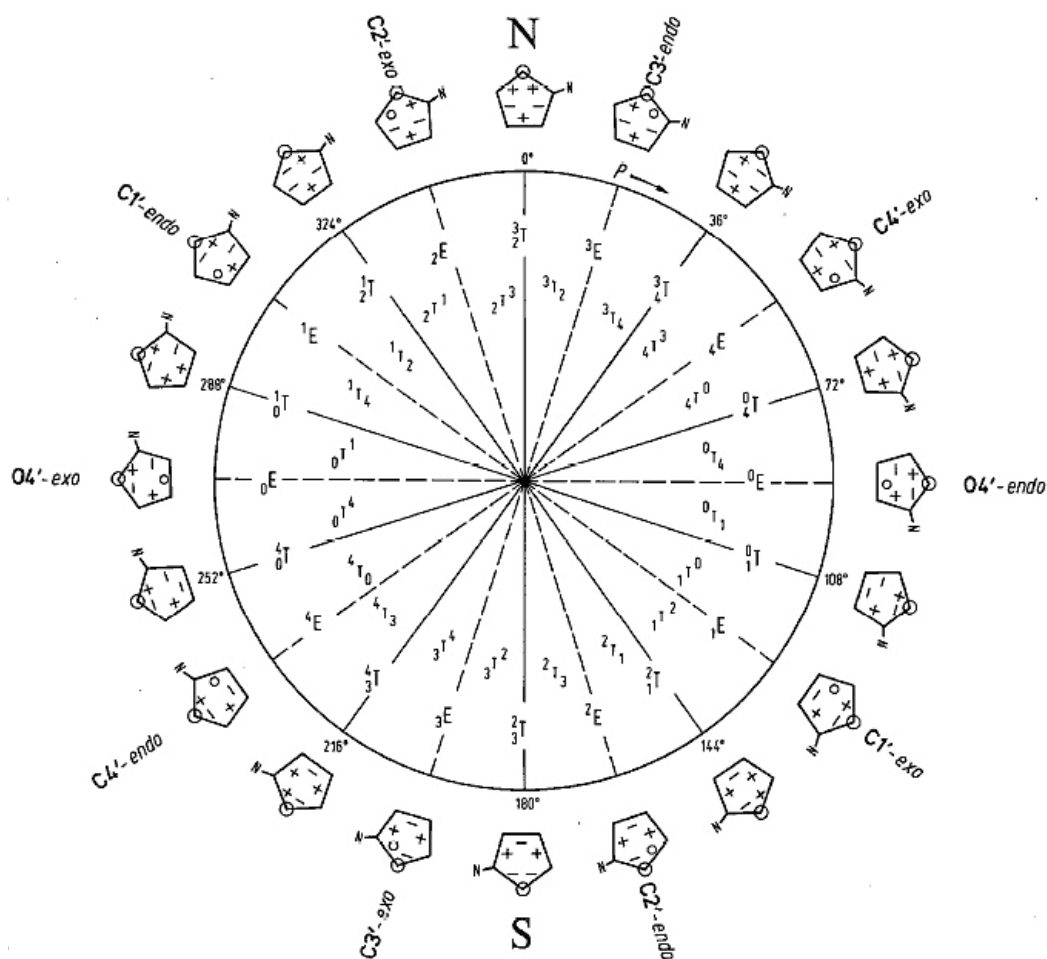


Figure 1.3 – Cercle de pseudorotation du cycle furanose pour les nucléosides. Les conformations dans la moitié supérieure du cercle sont appelées *Nord* (N) ($P = 0^\circ \pm 90^\circ$), celles dans la moitié inférieure sont appelées *Sud* (S) ($P = 180^\circ \pm 90^\circ$). Les conformations E se trouvent à des valeurs de phases multiples impaires de 18° tandis que les conformations T ont des valeurs de phase multiples paires de 18° (issue de [12]).

Les bases sont des molécules que l'on peut qualifier de planes. Les bases sont reliées au sucre par l'intermédiaire d'une liaison glycosidique entre un azote endocy- clique de la base et le carbone C1' du sucre. Pour les acides nucléiques naturels, la liaison glycosidique a une forme anomérique en β . De ce fait, les bases sont orien- tées du même côté que la liaison exocyclique C4'-C5' du sucre. La rotation des bases par rapport au sucre autour de la liaison glycosidique est décrite par l'angle de torsion χ (qui sera composé des atomes O4'-C1'-N9-C4 pour les purines et

O4'-C1'-N1-C2 pour les pyrimidines). La majorité des nucléotides a une conformation de l'angle χ dite *anti* (entre 180° et 300°), tandis qu'une minorité a une conformation *syn* (entre 30° et 90°).

Le groupement phosphate est arrangé en tétraèdre déformé. Le plus souvent, les liaisons O-P-O-C ont des angles de torsion avec des valeurs principalement autour de 60° et 300° . Dus à des effets stériques, conséquence de la présence du nucléoside, les liaisons O3'-P et P-O5' ont des angles de torsions typiquement aux alentours de 300° .

La chaîne nucléotidique a un espace conformationnel défini par 6 angles de torsion. Ces angles sont respectivement dans le sens 5'-3' α (O3'-P-O5'-C5'), β (P-O5'-C5'-C4'), γ (O5'-C5'-C4'-C3'), δ (C5'-C4'-C3'-O3'), ε (C4'-C3'-O3'-P) et ζ (C3'-O3'-P-O5') (Figure 1.4), ainsi que l'angle de torsion glycosidique défini précédemment.

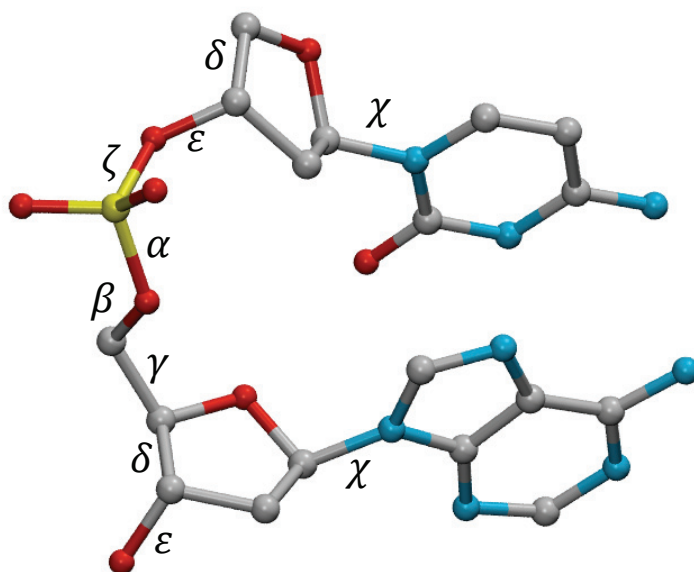


Figure 1.4 – Représentation des angles de torsions du squelette nucléotidique

La présence de ces sept angles de torsions définissant le squelette phosphodiester laisse présager un grand nombre de conformations potentielles des nucléotides. En réalité, même si ce squelette est un objet relativement flexible, le nombre

de conformations est limité par des contraintes stériques, restreignant ainsi le nombre de conformations accessibles. De ce fait, les six angles de torsions de la chaîne nucléotidique vont essentiellement avoir des conformations situées dans trois domaines prenant les valeurs de 60° , 180° et -60° nommés respectivement *gauche +* (g^+), *trans* (t) et *gauche -* (g^-) [13]. Ces restrictions stériques vont également entraîner des couplages entre différents paramètres du squelette.

Couplages conformationnels entre différents paramètres. Les paramètres décrivant le squelette nucléotidique étant tous reliés, un changement conformationnel de l'un d'entre eux pourra donc entraîner des changements à différents niveaux du squelette, ce qui est à l'origine du polymorphisme du squelette phosphodiester.

Parmi les différents couplages, les angles ε et ζ sont fortement corrélés. En effet, dans un ADN-B canonique, le couple ε/ζ se trouvera sous la forme t/g^- qui est communément reconnue comme le sous-état « BI ». Ses angles auront tendance à passer dans des conformations g^-/t nommé « BII », conformations qui seront en équilibre dynamique avec la première.

Un autre couplage important concerne les angles α/γ qui pourront transiter entre plusieurs conformations, les plus communs étant g^-/g^+ et t/t .

1.2.2 Paramètres hélicoïdaux

Les paires de bases Watson-Crick (WC) sont connectées par deux ou trois liaisons hydrogène relativement flexibles, leur permettant des déformations significatives de leur arrangement optimal. Depuis le début des années 80 et l'arrivée des premiers cristaux d'ADN [14], est apparue la nécessité de standardiser les définitions des différents paramètres hélicoïdaux afin d'optimiser les analyses. Ce fut chose faite suite à un colloque à Cambridge en 1989 [15], complété par la convention de Tsukuba dix ans plus tard [16].

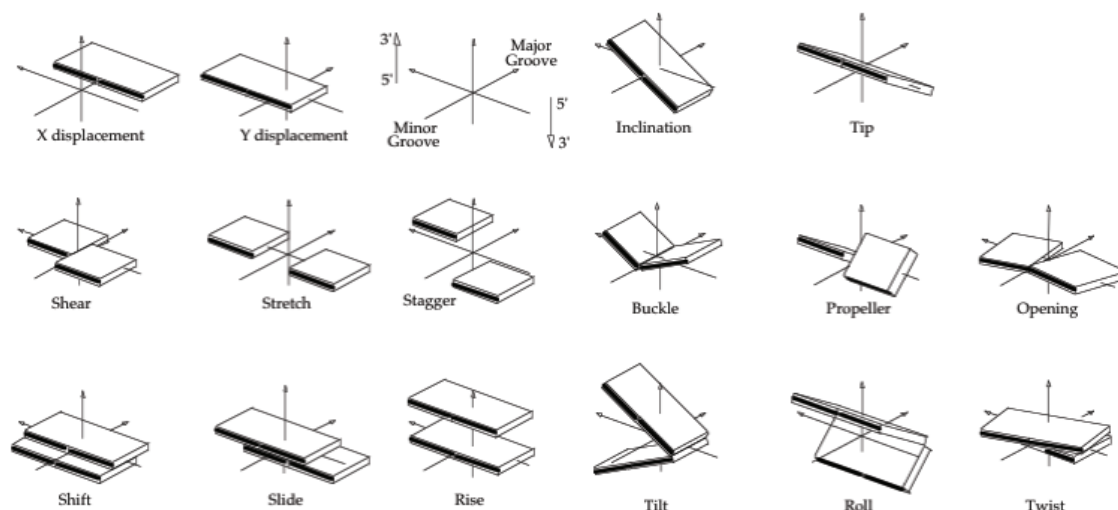


Figure 1.5 – Représentations schématiques des paramètres hélicoïdaux de l'ADN

Des paramètres décrivant la translation et la rotation ont donc été standardisés permettant ainsi de décrire la géométrie de paires de bases isolées (ou paramètres « *intra paires de bases* ») : *shear*, *stretch*, *stagger*, *buckle*, *propeller* et *opening*; la géométrie de deux paires de bases successives (ou paramètres « *inter paires de bases* ») : *shift*, *slide*, *rise*, *tilt*, *roll* et *twist* ; ainsi que la géométrie des bases par rapport à l'axe : *X-displacement*, *Y-displacement*, *inclinaison* et *tip* (Figure 1.5). Ces paramètres seront détaillés plus loin.

1.2.3 Les sillons de l'ADN

Les sillons de l'ADN sont les principaux sites d'interactions de toutes sortes de molécules allant de la petite molécule d'intérêt pharmaceutique (type molécule anti-cancer e.g. les benzodiazepines, la distamycine A ou encore le DAPI), interactions qui vont se faire via des liaisons avec le petit sillon [17], jusqu'à des grosses protéines telles que les facteurs de transcriptions [18], qui elles, auront plus tendance à se lier au niveau du grand sillon de l'ADN, voire parfois des protéines qui auront des interactions avec les deux sillons [19].

Le processus de reconnaissance en plus de faire intervenir des interactions locales, va concerner tous les paramètres structuraux et dynamiques des partenaires [20]. De ce fait, la conformation et la déformabilité de l'ADN, et plus particulièrement des sillons, a une grande influence sur la reconnaissance de sites spécifiques par des protéines d'interactions. Deux facteurs principaux vont influencer la géométrie des sillons : la conformation globale de l'ADN et les effets locaux de la séquence nucléotidique.

La conformation de l'ADN va donc influencer la géométrie des sillons. En effet, la forme A se caractérisera par un petit sillon large mais peu profond et un grand sillon profond et étroit, alors que la forme B (qui est la conformation de l'ADN en milieu physiologique) se caractérisera par un petit sillon étroit, un grand sillon large, mais des profondeurs plus ou moins comparables (voir paragraphe 1.2.4 pour plus de détails). Les différences de géométrie entre les sillons de l'ADN-B contribuent clairement à la spécificité de liaison à l'ADN. Le grand sillon semble être le plus important concernant les reconnaissances site spécifique, bien que quelques contacts ont lieu au niveau du petit sillon [18]. En effet, puisque le grand sillon est plus large, il sera donc plus accessible pour de potentielles liaisons hydrogène. L'analyse des motifs de donneurs et accepteurs réalisée par Seeman *et al.* [21] en 1976 a permis de mettre en évidence un mécanisme majeur de discrimination et de reconnaissance des paires de bases individuelles, à savoir la complémentarité des liaisons hydrogène au niveau des sillons de la double hélice. Cette reconnaissance pouvant se faire de manière directe, ou indirecte via une molécule d'eau [22]. La spécificité des liaisons hydrogène est plus grande au niveau du grand sillon puisque les quatre assortiments de paires de bases possibles ont chacun un motif unique de donneurs et accepteurs de liaison hydrogène contrairement au petit sillon où aucune discrimination ne pourra être faite entre les paires de bases A•T et T•A ainsi qu'entre les paires de bases G•C et C•G (voir Figure 1.6). La reconnaissance des motifs de donneurs et accepteurs au niveau des sillons par des protéines d'interactions est appelée « *direct recognition* » ou reconnaissance directe.

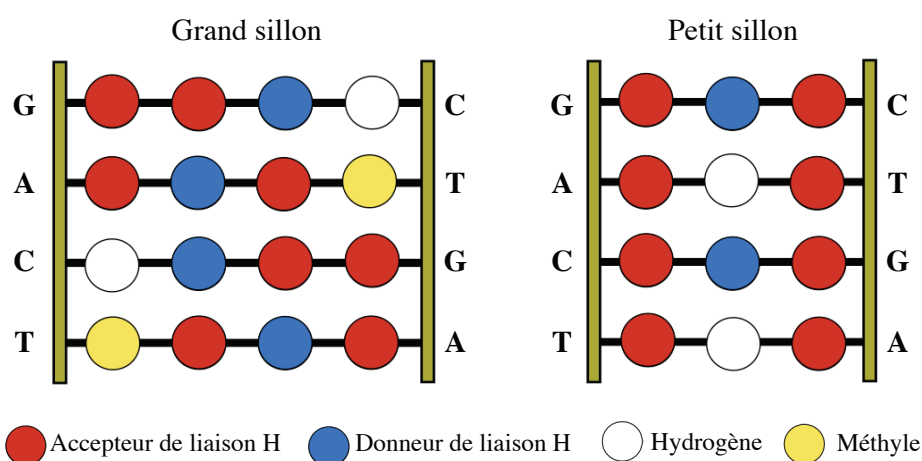


Figure 1.6 – Motifs des sites de donneurs et accepteurs de liaisons hydrogène dans les sillons.

La séquence nucléotidique a également un rôle prépondérant sur l'aspect des sillons. En effet, certaines séquences nucléotidiques auront tendance à induire des déformations de la géométrie des sillons. On pourra notamment citer des rétrécissements du petit sillon dus notamment aux séquences *A-tracts* (ou séquences *oligo-A*) [23-27], ou au pas purine – pyrimidine (RpY) [28], ou encore des rétrécissements du grand sillon dus eux au pas YpR [28]. Les déformations de la double hélice induite par la séquence seront développées dans une autre partie (voir 1.3).

1.2.4 Conformations de l'ADN

Depuis le modèle de Watson et Crick en 1953 [9], les multiples travaux qui suivirent permirent d'affiner le modèle de double hélice proposé ainsi que de fournir une vision précise de sa conformation. Il est très vite apparu que l'ADN pouvait adopter plusieurs conformations en solution sous différentes conditions expérimentales [29]. Trois conformations principales ont pu être définies ; deux conformations avec une hélice *droite* (ADN-A et ADN-B) et une conformation avec une hélice *gauche* (ADN-Z). Les doubles hélices antiparallèles avec une orientation

droite sont définies comme les formes *canoniques* de l'ADN. En effet, ce sont les formes principales adoptées par l'ADN dans les conditions du vivant [29].

D'un point de vue général, grâce à l'organisation en double hélice antiparallèle des deux brins d'ADN, les constituants des nucléotides pourront adopter des conformations optimales leur permettant de créer des liaisons hydrogène de types Watson-Crick entre les bases complémentaires et ainsi protéger leurs groupements hydrophobes du solvant environnant. La largeur des complexes formés par les paires de bases est quasi similaire pour les deux appariements A•T et C•G, ce qui permettra un empilement des paires de bases le long de la double hélice. La stabilité de la structure sera donc assurée d'une part, par les liaisons hydrogène et d'autre part, par l'énergie d'empilement des bases.

Hélices droites canoniques, formes B et A : Les structures de ces deux conformations ont pu être déterminées par des techniques de diffraction des rayons X sur des fibres d'ADN naturelles [30-34] supportées par des études comparatives basées sur des cristaux uniques [35].

L'ADN-B est la conformation la plus stable de la double hélice généralement favorisée en solutions aqueuses et en conditions physiologiques [36] (Figure 1.7). Cette conformation est donc constituée d'une double hélice dextrogyre avec des empilements parallèles de paires de bases WC (A•T et C•G) perpendiculaires à l'axe hélicoïdal. Le diamètre de cette hélice est d'approximativement 20Å. Chaque tour complet de l'ADN est composé d'environ 10 paires avec les paires de bases séparées par une distance de 3.4 Å avec un *twist* d'environ 36°. Sous cette forme, les désoxyriboses auront principalement des plissements C2'-*endo* avec un angle de torsion de la liaison glycosidique χ en *anti* (entre 240° et 270°). L'ADN-B idéal est caractérisé par un grand sillon large et un petit sillon étroit, tous deux d'une profondeur équivalente. En général, le grand sillon présentera des bases avec un nombre plus important de substituants (en O6 et N6 pour les purines et en N4 et O4 pour les pyrimidines), ce qui, combiné aux différences de largeurs des sillons, aura une importance capitale pour les interactions avec les autres molécules.

L'ADN-A est une conformation de la double hélice que l'on retrouvera dans des conditions d'hydratation plus faibles comme par exemple dans un environnement éthanol-eau [37,38] ainsi que dans quelques complexes ADN-protéines [38] (Figure 1.7). Comme la conformation B, l'ADN-A est formé d'une double hélice dextrogyre. Les bases quant à elles, sont excentrées par rapport à l'axe de l'hélice (d'environ 5 Å) avec une inclinaison (ou *tilt*) d'environ 20°, ainsi qu'un écartement inférieur d'environ 2,6 Å (mesuré le long de l'axe hélicoïdal) engendrant une augmentation du diamètre de l'hélice ainsi que du nombre de bases par tour (passant de 10 à 10,7). Dans cette conformation, les sucres présentent des plissements en C3'-*endo* et des angles de torsions χ d'environ 200°. Ces changements de paramètres hélicoïdaux comparés à l'ADN-B auront pour conséquences de changer la géométrie des sillons. De ce fait, l'ADN-A se caractérise par un grand sillon étroit et profond, tandis que le petit sillon est large et superficiel.

Hélice gauche, forme Z : Jusqu'à la fin des années 70, toutes les structures déterminées par des techniques de diffraction des rayons X de fibres ont été modélisées comme des doubles hélices droites. C'est donc avec surprise que la première structure résolue par cristallographie s'avéra avoir une double hélice « gauche » avec un squelette en zigzag [39,40], et fut en conséquence nommée ADN-Z (Figure 1.7). L'ADN prend cette conformation sous certaines conditions bien spécifiques qui sont une alternance de bases purines et pyrimidines (communément GpC) dans un environnement très concentré en sel (> 4 M NaCl) [41]. Dans cette conformation, les paires de bases sont inversées (*i.e.* retournées par rapport à l'ADN canonique) ; les sucres quant à eux, sont caractérisés par une alternance de plissement C2'-*endo* et C3'-*endo*. Les purines prennent une orientation *syn*. Ces particularités structurales vont avoir pour conséquence de générer un grand sillon peu prononcé et un petit sillon proche de celui de l'ADN-B mais avec un squelette phosphaté en zigzag. L'ADN-Z semble jouer un rôle dans la stimulation de la transcription [42] ainsi que dans des phénomènes de mutation [43].

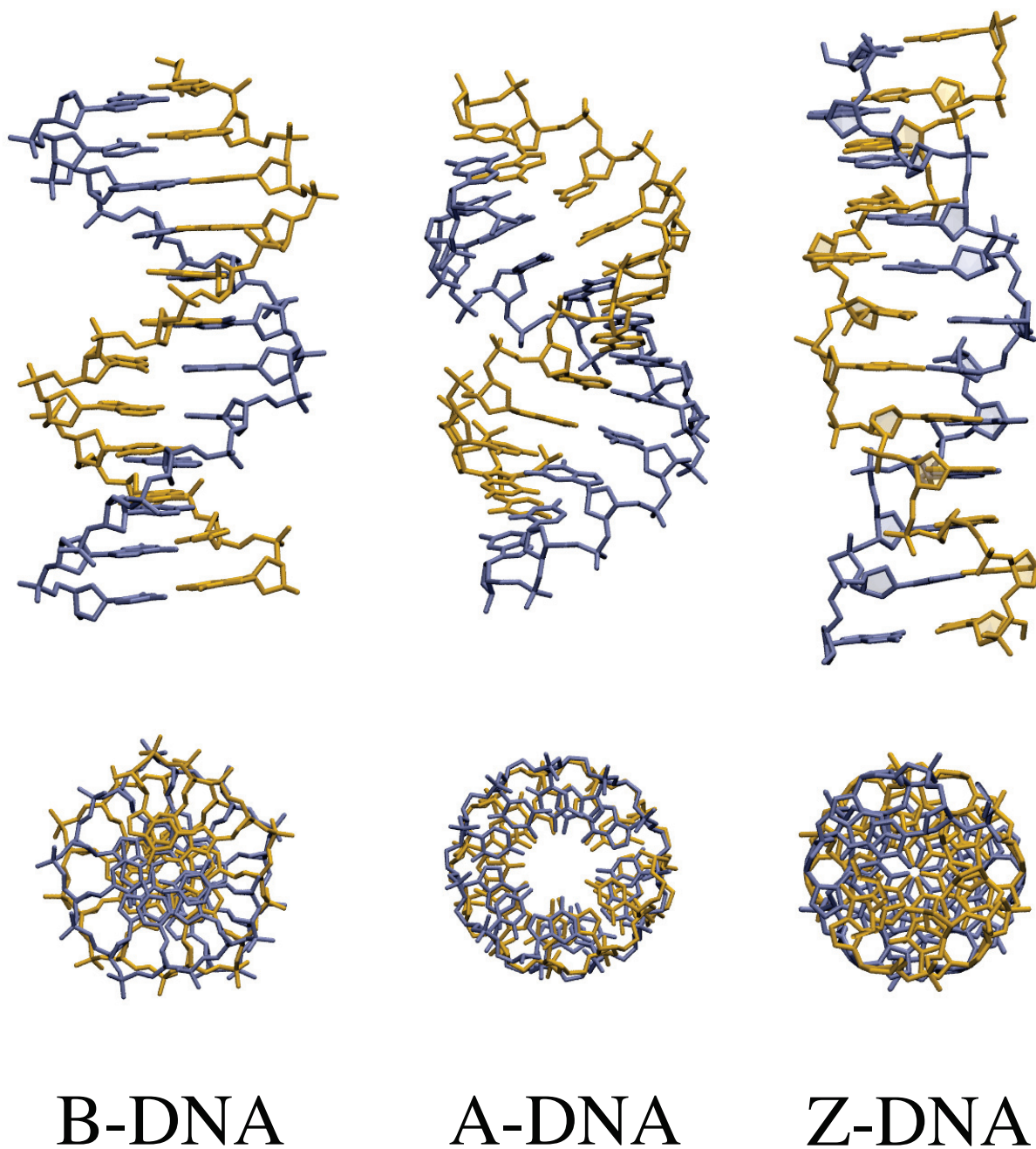


Figure 1.7 – Représentation des différentes conformations de l'ADN.

1.3 Les effets de séquences

Grâce aux nombreux travaux expérimentaux faisant suite au modèle de Watson et Crick parmi lesquels on pourra citer les travaux de Dickerson [14,44], l'image d'une double hélice isotrope a laissé peu à peu la place à une vision plus irrégulière de la structure d'ADN [45]. Il est maintenant clair que la reconnaissance de séquence spécifique par les protéines ne se fait pas seulement via des contacts entre les bases et les acides aminés mais aussi indirectement et que la séquence nucléotidique est à l'origine de variations de la double hélice. De ce fait, les variations séquences-spécifiques de la géométrie de la double hélice jouent donc un rôle essentiel dans les processus de reconnaissance protéine-ADN. Ce mode de reconnaissance est assimilé à un mode nommé « *indirect recognition* » ou reconnaissance indirecte [46-48] (en parallèle au mécanisme « direct » par lequel la reconnaissance entre les acides aminés et les bases se fait via les motifs de donneur et accepteur de liaisons hydrogènes des sillons (voir plus haut)).

Depuis de nombreuses années, de multiples données expérimentales concernant des complexes protéines-ADN ont pu être accumulées, ce qui nous a permis d'amasser une quantité d'informations sur les interactions spécifiques entre les protéines et les bases de l'ADN.

Un des problèmes émergeant de ces données est le manque de certitudes concernant ces complexes. En effet, dans les complexes protéines-ADN, on ne sait pas si les déformations de l'ADN sont intrinsèques à la séquence, ou si elles sont dues à l'interaction avec le partenaire.

Malgré des efforts soutenus par l'intermédiaire de diverses techniques telles que la cristallographie [28,49-51] ou encore la RMN [52-55], relativement peu de données sont disponibles concernant les effets de séquences de l'ADN libre, conséquence de la limitation de ces techniques (par exemple, en cristallographie le problème est dû au fait qu'en raison de sa forte flexibilité il est plus difficile de cristal-

liser de l'ADN seul plutôt que complexé) combiné à un manque de compréhension de l'étendue du problème en lui même.

Néanmoins, malgré la quantité limitée de données expérimentales disponibles, et grâce, par la suite, à des travaux faisant appel à diverses techniques visant à compléter ces lacunes (parmi lesquelles la modélisation par mécanique moléculaire, ou encore des simulations de Monte Carlo ou plus récemment de la dynamique moléculaire), de nombreuses caractéristiques de l'ADN-B ont quand même pu être mises en lumière.

Ces travaux ont tenté d'expliquer les variabilités conformationnelles de l'ADN via une corrélation avec les pas de paires de bases. Dans ce sens, on pourra notamment citer les analyses concernant les A-tracts, ou séquences poly-A, déjà citées précédemment, qui sont constituées d'un enchainement d'au moins trois dinucléotides ApA, TpT, ou ApT [26] consécutifs. Ces A-tracts engendrent un rétrécissement du petit sillon qui serait dû aux interactions entre les groupes fonctionnels adjacents des paires de bases dans le grand sillon [56-58] ainsi qu'une courbure locale de l'ADN, courbure pour laquelle différents modèles ont été proposés [59,60]. Les pas TpA quant à eux, présentent un empilement minimal résultant en une contrainte structurelle moins grande ainsi qu'à une variabilité conformationnelle plus importante [60]. De manière plus générale, la courbure locale induite par les pas de paires de bases semblerait suivre une règle très simple dictée par le modèle de « *steric clash* » proposé par Calladine en 1982 [49] qui sera affiné par la suite [61] : les dimères RpY et ApA auraient une courbure en direction du petit sillon, tandis que les pas YpR ainsi que les pas GpG auraient une courbure préférentielle dans le sens du grand sillon, les pas pyrimidine-purine (YpR) étant les pas ayant le plus d'anisotropie et de flexibilité [28,61,62]. Les analyses détaillées réalisées par Olson et ses collaborateurs [28,63], en plus de proposer une description des propriétés conformationnelles de l'ADN en terme de dimères de paires de bases, ont permis de poser les bases des champs de forces pour l'ADN [28].

Cependant, d'autres travaux sont venus par la suite montrer que les dinucléotides sont affectés par les paires de bases adjacentes [51,64,65] et donc que la des-

cription des conformations de l'ADN via les dinucléotides semble insuffisante. De ce fait, une question vient à émerger : si un modèle basé au niveau du dinucléotide et l'utilisation de 10 fragments est insuffisant qu'en sera t'il d'un modèle à l'échelle du trinucleotide [66] qui lui impliquera 32 fragments, ou encore du tetranucleotide impliquant quant à lui 136 fragments ? Dans le cas du modèle tetranucleotide, seul environ une centaine des 136 tetranucleotides possibles sont actuellement cristallisés (dû notamment au faible nombre de cristaux d'ADN-B libre disponibles ainsi qu'au fait que certaines séquences sont plus difficiles à cristalliser que d'autres). De plus, une autre limitation est amenée par le fait que les conditions de cristallisation pourraient affecter la géométrie de certains tetranucleotides et donc, il sera difficile d'extraire des données fiables de structures dont un seul cristal est disponible [67].

En conséquence, l'utilisation de la dynamique moléculaire semble être la réponse la plus adaptée pour compléter le manque de connaissance des effets de séquences. En effet, grâce à des avancées sur différents points (points que nous détaillerons plus tard), l'application de la dynamique moléculaire à l'étude de l'ADN a fait de nombreuses avancées ces dernières années [68]. Dans ce sens des études visant à compléter nos lacunes ont été commencées telles que celles menées par Sarai [69] et ses collaborateurs, ou encore celle de Gardiner *et al*, qui a permis de démontrer notamment que l'étude au niveau des octa-nucléotides est trop élevée, et qu'une distinction d'effets spécifiquement liés à une séquence unique n'est plus possible [70,71]. On pourra également citer les travaux proposés par l'« *Ascona B-DNA Consortium* » (ABC) [72-74] qui en plus d'amener une correction [75] au champ de force qui est aujourd'hui le plus adapté à la description de l'ADN [76], arrive à la conclusion que l'étude de l'effet de séquence au niveau du tetranucleotide semble encore insuffisante et qu'il faudrait sans doute envisager une étude à l'échelle du penta-nucléotide, voire peut-être de l'hexa-nucléotide [74]. Pour l'instant, il est donc toujours impossible de prédire avec précision les effets de séquence d'une chaîne nucléotidique, et ce quels que soient les niveaux auxquels on considère la double hélice d'ADN.

1.4 Propriétés dynamiques et mécaniques de l'ADN

Afin de réaliser les interactions entre les protéines et l'ADN, un grand nombre d'adaptations structurelles de l'un ou l'autre partenaires, voire les deux, peuvent être nécessaires. Ces adaptations sont donc intimement liées à la dynamique des molécules, c'est pourquoi il est important d'obtenir le plus d'information possible sur la structure, la dynamique ainsi que l'environnement de ces différentes molécules et tout particulièrement de l'ADN.

La double hélice d'ADN possède un polymorphisme conformationnel. Notamment, son squelette phosphaté transite rapidement entre différents sous états. Il est aujourd'hui également clair que la présence d'ions dans le milieu est un élément essentiel à la conformation et à la stabilité de la double hélice d'ADN [77]. Puisque l'ADN et les autres molécules évoluent dans un milieu physiologique, le rôle de l'eau et des ions entourant la double hélice d'ADN ainsi que leurs influences sur les conformations de l'ADN ne peuvent donc pas être ignorés.

1.4.1 De l'eau et des ions

La molécule d'ADN est un poly-électrolyte fortement anionique. De ce fait, de nombreuses interactions sont donc possibles avec son environnement [78,79]. Il est reconnu que la stabilité et la flexibilité des nucléotides sont grandement influencées par les interactions avec les molécules d'eau et les ions les entourant [80]. L'eau et les ions vont non seulement avoir des interactions avec l'extérieur de la double hélice en neutralisant les charges portées par les groupements phosphates, mais également avec l'intérieur de la structure dans les sillons.

L'hydratation de la double hélice est indispensable à sa conformation et à ses propriétés à un tel point que l'eau est reconnue depuis les années 80 comme faisant partie intégrante de la structure des acides nucléiques [81-84].

La molécule d'ADN est entourée de couches d'eau étroitement liées à sa surface [85,86], c'est pourquoi l'étude des différentes couches est une clé de la compréhension de son polymorphisme et des mécanismes de reconnaissance spécifique. De nombreuses études utilisant différentes techniques expérimentales se sont penchées sur la question de l'hydratation, la cristallographie s'avérant être la méthode de choix pour étudier la première couche d'hydratation [77].

Les premières études aux rayons X de fibres d'ADN, ont permis de mettre en évidence que les conformations A et B de l'ADN étaient dépendantes du taux d'humidité. En effet, si l'activité de l'eau est faible, l'ADN se trouvera sous la forme A, alors que si l'activité est forte, elle prendra la forme B [87]. Par la suite, la mesure de l'hydratation de l'ADN a pu être affinée par le biais de différentes approches expérimentales.

Grâce notamment à la résolution de la structure d'ADN et aux nombreuses études cristallographiques qui suivirent, l'importance des couches de solvation autour de la double hélice d'ADN a pu être mise en évidence. Une bonne description des différentes couches de solvant entourant l'ADN s'avère nécessaire pour la bonne compréhension des mécanismes influençant les comportements de la double hélice.

L'hydratation des paires de bases WC a pu être définie grâce notamment à des analyses cristallographiques de poly-nucléotides par Schneider et ses collaborateurs [88-91]. Grâce à cette série d'analyses systématiques de la double hélice d'ADN, ils ont pu montrer que les molécules d'eau sont organisées autour des bases [89,90] et des phosphates [91], et que les motifs d'hydratation dépendent aussi bien de la conformation de l'ADN (l'hydratation des formes A et B étant comparables entre elles, mais distinctes de la forme Z) que du type de base (Figure 1.8). Les auteurs ont également conclu que les molécules d'eau se regroupent près des atomes hydrophiles formant des « *cluster* » autour des acides nucléiques. Ces résultats leurs ont permis d'introduire le concept d'« *hydrated building blocks* » qui sont des blocs de bases d'acides nucléiques avec le site d'hydratation qui leur est associé.

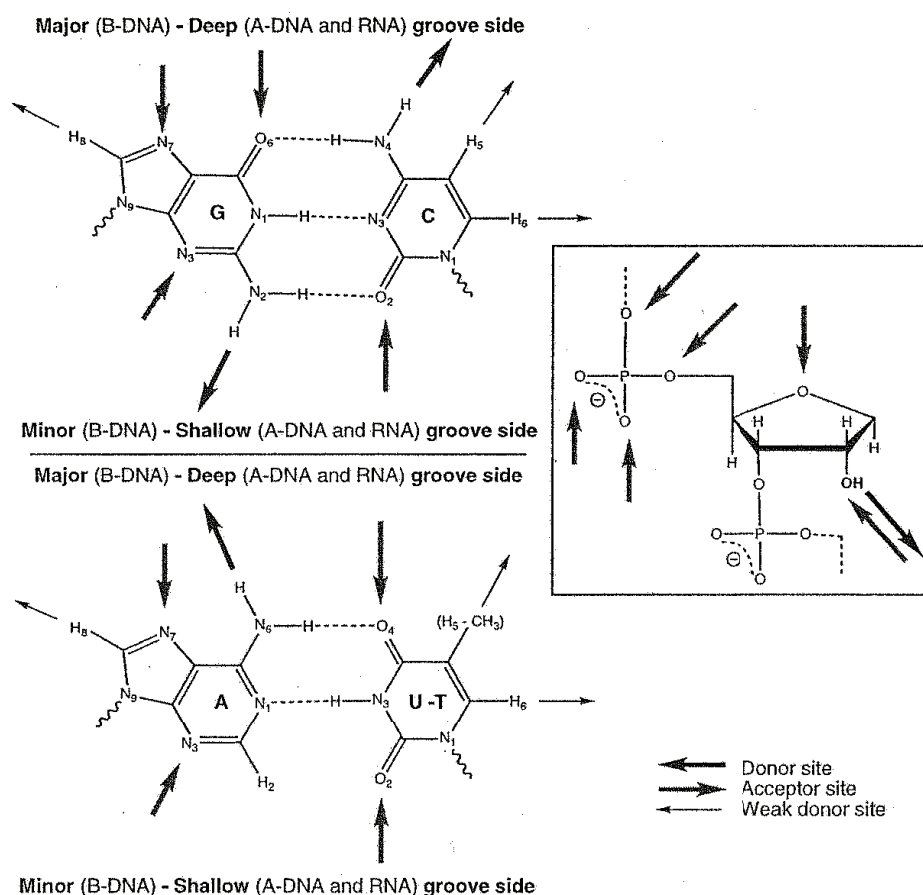


Figure 1.8 – Sites d'hydratation potentiels autour des paires de bases WC et du squelette sucre phosphate (issu de [80]).

Les études cristallographiques des premières structures de la double hélice ont également permis de montrer que l'hydratation des sillons de l'ADN-B en plus d'être différente selon le sillon est également essentielle à sa structure [92,93]. En effet, dans le grand sillon une unique couche d'eau partiellement ordonnée semble être présente, tandis que dans le petit sillon, deux couches d'eau paraissent organisées en « *spines of hydration* » (définie pour la première fois grâce à l'analyse de la structure cristalline d'ADN-B de séquence 5'- CGCGAATTCGCG-3' déterminée par Drew et Dickerson [92]) que l'on pourra traduire comme « épine dorsale d'hydratation » qui comme son nom l'indique, sera essentielle à la structure de l'ADN. Cette épine dorsale d'hydratation en zigzag est séquence dépendante

[92,94-97] ; différents motifs d'hydratation ont notamment été observés pour des séquences avec des petits sillons élargis [98,99].

Beaucoup de techniques expérimentales sont incomplètes. En effet, bien que les analyses des structures par cristallographie de rayons X nous aient fourni de nombreuses informations sur les états d'hydratation de l'ADN [100], ces informations restent partielles puisque cette méthode généralement ne nous permet pas de voir les hydrogènes, ni d'avoir une vision de l'hydratation à concentration physiologique ce qui pourra avoir une influence sur les structures hydratées [101], ou plus simplement, ne nous permettra pas d'avoir une vision complète des couches d'hydratation par le fait qu'il est rare de détecter toutes les molécules d'eau entourant les nucléotides. Les techniques de RMN quant à elles, nous permettront généralement de détecter uniquement les molécules d'eau à proximité d'un proton non échangeable couplé à un temps de résidence long. Face à la limitation des analyses structurales de l'hydratation de l'ADN, l'utilisation de méthodes computationnelles semble une alternative adaptée pour compléter les analyses expérimentales.

Des études par dynamiques moléculaires sont notamment venues souligner encore davantage la différence d'hydratation des sillons en mettant en avant qu'une densité plus importante de molécules d'eau était localisée dans le petit sillon tandis que dans le grand sillon l'hydratation est inférieure révélant une plus large mobilité et donc un temps de résidence réduit pour les molécules d'eau interagissant avec le grand sillon [102-105]. Il est nécessaire de garder en tête que cette technique est dépendante des approximations des champs de forces. Malgré ceci, les résultats obtenus par dynamique moléculaire sont cohérents avec ceux obtenus par techniques expérimentales, et sont capables de générer des distributions réalistes des molécules d'eau autour des nucléotides ainsi que d'avoir une vision dynamique de l'ensemble [104,106].

Bien qu'il y ait clairement une corrélation entre les conformations de l'ADN, le rôle exact des molécules d'eau dans les transitions de conformations de l'ADN n'est pas encore tout à fait compris. Cependant, l'eau n'est pas l'unique facteur pouvant

influencer la structure de l'ADN. Les ions présents dans le milieu vont également avoir un rôle à jouer dans le polymorphisme de la double hélice d'ADN.

Les contre-ions, au même titre que les couches d'hydratation, sont essentiels à la stabilité de la double hélice, et par conséquent au rôle biologique des acides nucléiques. L'ADN étant un poly-électrolyte anionique (dû à son squelette phosphaté), il aura donc de fortes interactions avec les contre-ions cationiques présents dans le milieu. Ces contre-ions vont notamment empêcher les deux brins de la double hélice de se dissocier sous la force de répulsion des groupes phosphates en les neutralisant. Si à la fin des années 70, les premiers modèles représentaient les contre-ions comme un nuage séquence-indépendant autour de la double hélice [107], il est aujourd'hui clairement établi, tant sur un point expérimental que par simulation, que les ions (au même titre que les molécules d'eau) auront un rôle structural bien plus important en occupant des sites préférentiels tels les alentours des groupements phosphates du squelette ainsi que les sillons [108-112].

Tout comme les couches d'hydratation, il est très difficile d'étudier les interactions ADN-cations par des méthodes expérimentales telles que la cristallographie de rayon X (notamment pour des limitations intrinsèques à la technique telles que les conditions de cristallisation, le cristal en lui-même qui pourra influencer la position des ions, ou encore la localisation des ions dans les densités électroniques qui reste très difficile). Cependant, la combinaison des techniques expérimentales avec la MD ont permis de mettre en lumière des évidences quant aux interactions entre l'ADN et les cations. Des études de la structure cristalline du dodécamère de Drew-Dickerson ont montré que les cations monovalents Na^+ et K^+ partiellement déshydratés se liaient au petit sillon de l'ADN [113,114], avec notamment le Na^+ pouvant pénétrer dans la portion centrale de l'épine dorsale d'hydratation du petit sillon [92,113]. D'autres études par spectroscopie RMN [108,115,116] et de MD [103,117] ont montré des interactions séquences spécifiques entre la double hélice d'ADN et les cations physiologiquement importants Na^+ , K^+ , Ca^{2+} et Mg^{2+} . Par exemple, en solution, les cations sont principalement localisés dans le grand sillon

au niveau de séquences riches en GC et dans le petit sillon au niveau de séquences riches en AT [112]. Ces liaisons cationiques séquences-dépendantes contribuent au polymorphisme conformationnel de l'ADN et se feraient plus précisément à la courbure de l'ADN et au rétrécissement de la largeur du petit sillon. En plus de leurs effets structuraux dépendants de la séquence, les ions ont un effet sur la stabilité globale de la double hélice. En effet, les ions présents dans le milieu vont influencer l'activité de l'eau [118] (une augmentation de la concentration en sel réduisant cette activité). Par conséquent, la propension de l'ADN à passer d'une conformation B à une conformation A et *vis versa*, sera donc liée à la concentration de sel [119].

De plus, la présence de contre-ions dans le milieu, dont notamment de K^+ et Na^+ via des interactions avec les phosphates, est à l'origine de la stabilisation de certains sous états du squelette phosphodiester [120].

1.4.2 Polymorphisme conformationnel et influence du milieu

La double hélice d'ADN et plus précisément son squelette phosphodiester présente un polymorphisme avec certains des angles le composant ayant des sous états qui transitent rapidement de l'un à l'autre.

Transitions BI/BII : Nous avons vu précédemment que les angles ε et ζ étaient couplés et qu'ils se trouvaient sous une forme canonique dans un sous-état appelé BI (t/g^-). Des études cristallographiques [13,121-123], confirmées par des études de spectroscopie RMN [124-126], ont révélé l'existence d'un deuxième sous état de ces deux angles nommés BII (g^-/t), les deux sous-états BI/BII pouvant s'interconvertir librement en solution. Par la suite, des études de dynamiques moléculaires ont permis d'établir que l'interconversion des conformations BI/BII, n'est pas un processus se faisant au hasard.

Du aux nombreuses interactions entre les groupements phosphates de l'ADN et des protéines, l'équilibre BI/BII qui va modifier la position des phosphates

pourrait donc avoir un rôle important dans les processus de reconnaissance indirecte de sites spécifiques [126-130], c'est pourquoi la compréhension de ces transitions et des règles qui les régissent est d'une importance non négligeable aujourd'hui.

Ces transitions sont réversibles à l'échelle de la nanoseconde même si dans certains cas elles peuvent être stabilisées au-delà d'une dizaine de nanosecondes [74]. Des études de modélisations moléculaires et de dynamiques moléculaires, ont montré que les interconversions BI/BII avaient une influence dans la courbure de la double hélice [13,131,132] ainsi que sur certains paramètres hélicoïdaux [74,131]. En 2007, une étude utilisant la spectroscopie RMN menée par Heddi et ses collaborateurs a permis de montrer que non seulement le ratio de conformations BI/BII de l'ADN-B affectait toute la structure de la double hélice, mais également que les populations BI/BII pouvaient être influencées par le type contre-ions monovalent (Na^+ ou K^+) interagissant avec le squelette phosphodiester [120]. Leurs observations ont également permis de conclure que les ions Na^+ et K^+ influencent tous deux (mais d'une façon différente) la structure de l'ADN.

D'autres études ont également tenté de démontrer que le dynamisme de transition BI/BII était lié à la séquence [64,73,74,131,133], mais pour l'instant, même si cette corrélation paraît évidente, aucune étude n'a encore pu en établir une règle bien définie.

Les angles de torsions α et γ du squelette phosphodiester de l'ADN, qui sont eux-aussi couplés entre eux, ont également été recensé comme pouvant potentiellement adopter des conformations non canoniques. En effet, en plus de la forme canonique que l'on retrouve dans la quasi totalité des cristaux d'oligomères d'ADN-B libre, dont les deux angles présentant des conformations g^-/g^+ pour respectivement α/γ , les cristaux de complexes ADN-B/protéines présentent souvent des formes non-canoniques de ces deux angles de torsion [134]. Ces formes non-canoniques ont pu être observées dans diverses études de dynamiques moléculaires [135-138], dont certaines ont proposé une corrélation de ces transitions avec la présence de contre-ions au niveau des phosphates [137]. Récemment, des travaux de dynamiques moléculaires menés dans le cadre de l'*Ascona B-DNA con-*

sortium, ont montré que les interconversions des formes canoniques vers des formes non-canoniques des angles α/γ se passaient dans une échelle de temps au-delà de la dizaine de nanosecondes. De façon plus importante, ces études ont également mis en évidence un mauvais traitement des angles de torsion α/γ dans les champs de forces utilisés en dynamiques moléculaires, entraînant une sur-stabilisation de formes non-canoniques de ces angles [72-74]. Les dynamiques moléculaires menées par la suite, intégrant une modification de ce champ de force (parmbsc0) [74,75], montrent que les oligomères d'ADN-B libres présentent des sous-états non-canoniques des angles α/γ dans moins de 1 % des simulations (avec des temps de transition supérieurs à dix nanosecondes).

Puisqu'elles sont présentes dans les complexes d'ADN-B/protéines, les transitions α/γ semblent donc clairement avoir un rôle dans les processus de reconnaissance entre les partenaires. Pour l'instant l'analyse de ces transitions est encore limitée (notamment en dynamique moléculaire) principalement par la fréquence de conversion entre ses sous-états (supérieur à une dizaine de nanosecondes).

CHAPITRE 2

Molécules intercalantes

2.1 Modes d'interactions avec l'ADN

La plupart des processus biologiques sont dépendants de la reconnaissance moléculaire ainsi que des interactions réversibles entre les molécules. Les acides nucléiques sont particulièrement concernés puisque tous les processus nécessaires au bon fonctionnement de la cellule nécessitent de nombreuses interactions. En plus des partenaires « classiques » tels les protéines ou autres nucléotides indispensables au bon fonctionnement de la cellule, de nombreuses petites molécules sont capables de former des complexes avec les acides nucléiques. De ce fait, ces petites molécules ont fait et font actuellement l'objet de quantités d'études et de développements dans le cadre de thérapies (telles les thérapies anticancer par exemple) [139,140]. Depuis les années 1960, il est reconnu que les interactions entre l'ADN-B et les molécules de faibles poids moléculaires présentes dans le milieu, qu'elles soient de type covalente ou non-covalente, pourront se faire selon trois modes principaux : par interactions avec le squelette sucre-phosphate anionique (Figure 2.1a) ; par intercalation entre deux paires de bases (Figure 2.1b) ; par des interactions avec les sillons de l'ADN (Figure 2.1c).

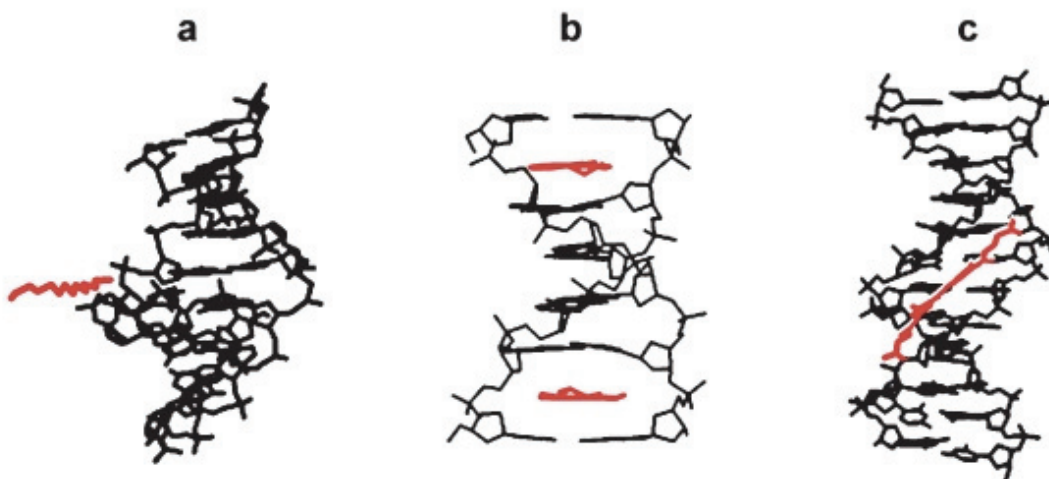


Figure 2.1 – Représentation des trois modes de liaisons principaux entre l'ADN et de petites molécules. **a.** Liaisons au squelette. **b.** Intercalation. **c.** Liaisons au sillon.

2.1.1 Modes d'interactions

2.1.1.1 Liaisons aux sillons de l'ADN

Dans les sillons, les ligands peuvent former des contacts directs avec les groupements fonctionnels du bord des bases que ce soit dans le grand ou le petit sillon. Ces liaisons pourront être covalentes ou non-covalentes.

Interactions avec le grand sillon

Le grand sillon donnera lieu essentiellement à des interactions avec des protéines ou des oligonucléotides d'origines naturelles, semi-synthétiques ou analogues. Les protéines auront tendance à se lier à l'ADN via des motifs en hélice α (motifs particulièrement adaptés au grand sillon de par leurs tailles et formes tels ceux contenus dans les domaines en *hélice-tour-hélice*).

La reconnaissance du grand sillon par des oligonucléotides nommés TFOs (pour *triplex-forming oligonucleotides*) [140-142] sera guidée par la création de liaisons hydrogène avec le bord des bases purines correspondant à des appariements *Hoogsteen direct* ou *Hoogsteen inverse*, formant ainsi des ADN triple brins nommés triplex. Ce type de reconnaissance séquence spécifique est utilisé dans la régulation des gènes.

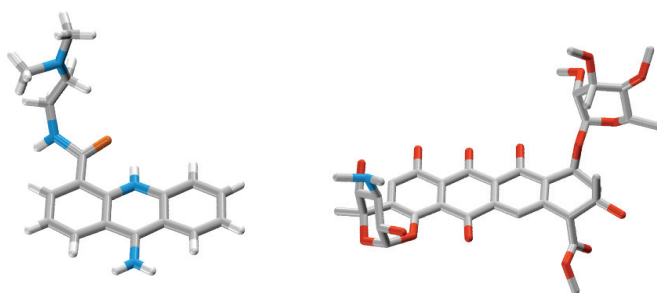


Figure 2.2– Représentation de molécules interagissant avec le grand sillon de l'ADN. Le DACA est représenté à gauche et la nogalamycine à droite.

Les molécules synthétiques ciblant le grand sillon sont rares, conséquence de leurs petites tailles les rendant plus à même d'effectuer des interactions via le petit sillon. On pourra cependant citer des molécules telles le DACA ou la nogalamycine (Figure 2.2).

Interactions avec le petit sillon

Ce mode d'interaction fut initialement proposé par Wells et ses collaborateurs en 1974 [143] grâce à leurs travaux sur la netropsine (Figure 2.3).

De nos jours, les petites molécules se liant au petit sillon sont l'une des classes parmi les plus étudiées, caractérisées par une haute spécificité de séquence. En plus d'être d'origines naturelles (telle la distamycine A), synthétiques (e.g. les diarylamidines parmi lesquels le DAPI ou Hoechst 33258) ou semi-synthétiques (telles les combilexines), ces molécules, dont la plupart se lieront à des séquences riches en bases thymine-adénine, présentent des propriétés très variées ; antiviral, antiprotozoaires, antibactériennes voire même antitumorales [17]. Certaines de ces molécules sont également utilisées comme marqueur dans les études de microscopie telles le DAPI ou encore Hoechst 33258 [17].

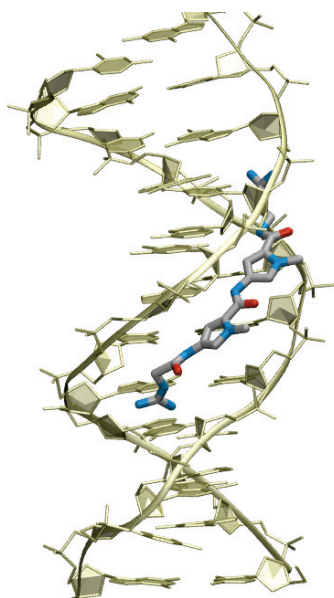


Figure 2.3 – Représentation de la netropsine liée au petit sillon de l'ADN (IDPDB : 1DNE).

Par conséquent, le petit sillon est devenu une cible thérapeutique intéressante en vue du développement de molécules d'activités biologiques.

Le petit sillon pourra également être reconnu par des protéines (même si le plus souvent elles reconnaîtront le grand sillon). Dans ce cas les liaisons se feront souvent via des structures en feuillets β . Ces interactions protéines/petit sillon seront à l'origine de courbures dans la double hélice liées à la déformation du sillon.

Liaisons covalentes aux bases :

Un autre type d'interactions se fera via des liaisons covalentes entre des petites molécules d'intérêt biologique et les bases de l'ADN. Ces liaisons se feront souvent sur l'atome N7 de bases puriques localisé dans le grand sillon, mais des interactions ont également lieu avec d'autres sites, comme par exemple avec le N2 de la guanine via le petit sillon dans le cas des dérivés de benzo[a]pyrène. Ce type de molécules, inclut des cancérogènes, mais également des molécules d'intérêt thérapeutiques, parmi lesquelles on trouvera le cisplatine et ses dérivées [144]. La liaison du cisplatine à l'ADN, impliquant une ou deux bases puriques (préférentiellement les guanines), pourra notamment être à l'origine de fortes courbures de la double hélice.

2.1.1.2 Liaisons au squelette sucre phosphate

Le squelette anionique de l'ADN est également reconnu comme un site potentiel de liaisons. La surface du squelette étant anionique et riche en oxygène, aura de ce fait, des interactions avec les cations métalliques alcalins et alcalino-terreux. La plupart des protéines auront également des interactions avec le squelette notamment via des résidus cationiques telles les arginines et les lysines. Cependant, peu de molécules synthétiques sont reconnues pour avoir des interactions à ce niveau. On pourra quand même citer les travaux de Farrel et de ses collaborateurs

[145] décrivant le triplatinNC (Figure 2.4) qui se complexe avec les oxygènes des groupements phosphates composant le squelette de l'ADN.

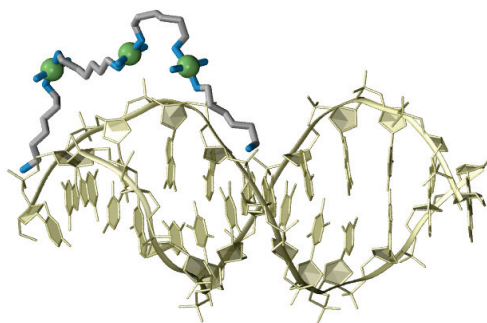


Figure 2.4 – Interactions entre le triplatinNC et le squelette phosphaté de l'ADN. Les platines étant représentés en vert (IDPDB :2DYW).

2.1.1.3 Intercalation entre deux paires de bases

Ce mode d'interaction qui fait partie de notre étude sera détaillé en profondeur dans le paragraphe suivant.

2.2 Intercalation et molécules intercalantes

Comme nous l'avons vu précédemment, les petits ligands peuvent non seulement interagir avec l'ADN en se liant aux sillons de l'ADN, avec les bases de l'ADN ou encore avec le squelette, mais également via un processus d'intercalation.

2.2.1 Intercalation « classique »

2.2.1.1 Principe

Décrite pour la première fois au début des années 1960 suite aux travaux de Lerman [146] l'intercalation consiste à l'insertion d'une petite molécule ou d'un fragment moléculaire entre deux paires de bases, formant une structure stable.

Une molécule intercalante classique pourra être représentée comme un ensemble de cycles aromatiques conjugués, désigné en tant que chromophore. Ces molécules sont composées généralement de deux à six cycles formant une plateforme d'une taille équivalente aux paires de bases de l'ADN. Certains intercalants simples, tels l'éthidium ou les proflavines, seront uniquement composés de ce chromophore qui sera généralement cationique à pH neutre. Cependant, la plupart des molécules intercalantes pourront avoir de nombreux substituants chimiques attachés à leurs chromophores, dont notamment des sucres ou des groupements peptides. Ces groupements fonctionnels additionnels, pourront avoir une influence sur l'orientation du ligand, sa spécificité de séquence, ainsi que sur la stabilité thermodynamique du complexe.

Le processus d'intercalation en lui-même implique l'ouverture de l'espace entre deux paires de bases empilées, et l'insertion du chromophore de façon parallèle aux paires de bases à partir du grand ou du petit sillon dans la poche hydrophobe résultante, formant ainsi des interactions π - π entre le chromophore et les bases au-dessus et au-dessous.

2.2.1.2 Déformations de l'ADN

La formation du complexe ligand/ADN qui est l'intercalation de molécules exogènes dans l'ADN aura des conséquences sur la structure de l'ADN.

En effet, pour permettre l'intercalation, les bases doivent se séparer d'environ 3.4 Å et ainsi former une cavité d'accueil pour le ligand. Cet écartement des bases

engendrera un déroulement local de la double hélice, dont l'amplitude sera dépendante de la géométrie et des propriétés physico-chimiques de la molécule se liant à l'ADN. L'angle de rotation entre deux paires de bases ("*twist*") de l'ADN-B est en moyenne 34° - 36° ; l'intercalation du bromure d'éthidium réduira cet angle de torsion de 10° , tandis que la liaison de proflavines le réduira d'environ 17° et celle d'anthracylines de 11° .

L'intercalation de petites molécules perturbera la structure de l'ADN à plus ou moins longue distance. En effet, en plus de modifier l'angle de rotation entre des bases adjacentes du site d'intercalation, compte tenu de la rigidité en twist de l'ADN, ce désenroulement local sera propagé le long de la structure du double brin.

Le processus de séparation des bases se produisant durant l'intercalation sera également à l'origine de nombreux changements au niveau du squelette phosphodiester ou de la géométrie des paires de bases (comme notamment l'éthidium pouvant influencer la courbure des paires de bases au niveau du site d'intercalation).

2.2.1.3 Le principe d'exclusion du plus proche voisin

Le phénomène d'intercalation donnera également lieu au « principe d'exclusion du plus proche voisin » (ou « nearest neighbour exclusion principle »).

Lorsque l'ADN est saturé avec une molécule intercalante simple, on pourrait s'attendre à ce que chaque site d'intercalation potentiel soit occupé (soit une stœchiométrie de un intercalant pour une paire de bases). Cependant les résultats expérimentaux ont montré une stœchiométrie d'un intercalant pour deux paires de bases à saturation en intercalant [147,148]. Chothers [149] explique ce résultat en proposant que chaque second site (voisin le plus proche) suivant un site d'intercalation le long de la double hélice demeure inoccupé. Ce qui en d'autres termes voudra dire qu'à chaque fois qu'un intercalant viendra occuper un site, la liaison d'un ligand supplémentaire sur le site immédiatement adjacent sera inhibé.

Ce principe communément admis sera toutefois remis en cause dans le cas de l'ADN-I. Cette conformation de la double hélice décrite pour la première fois en 1993 [150,151] présente la particularité d'une intercalation systématique, avec

une alternance de paires de bases orthogonales les unes aux autres, appartenant à deux doubles hélices formées de paires de cytidines hémiprotonées ($C\bullet CH^+$). Il faut cependant noter que l'ADN-I est constitué de deux duplexes parallèles qui peuvent avoir des propriétés différentes de la double hélice antiparallèle de l'ADN canonique.

2.2.2 Autres types d'intercalants et d'intercalations

2.2.2.1 Bis-intercalation

Les *bis-intercalants* (Figure 2.5) sont des molécules bifonctionnelles qui sont formées de deux systèmes plans aromatiques reliés par des chaînes de longueurs variables. De telles molécules seront généralement des ligands ayant une plus haute affinité, et une constante de dissociation plus faible que les mono-intercalants équivalents.

La constante de liaison d'un bis-intercalant sera équivalente approximativement au carré de celle du monomère. L'activité biologique étant reliée aux constantes d'associations, ce type de molécules aura de potentielles applications thérapeutiques très intéressantes. De plus, ce type d'intercalant se liant avec plusieurs paires de bases simultanément, aura donc un site de liaison plus étendu que celui des monomères, ce qui permet d'augmenter sa spécificité.

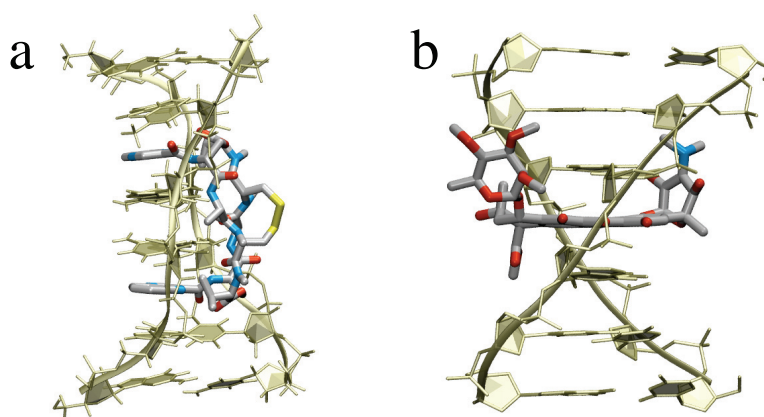


Figure 2.5 – **(a)** Représentation d’une interaction entre un bis-intercalant et l’ADN en l’occurrence une triestinA (IDPDB : 185D). **(b)** Représentation d’une intercalation enfilée entre la nogalamycine et l’ADN (IDPDB : 282D).

2.2.2.2 Threading intercalation

Ce type d’intercalation que l’on pourra traduire comme « *intercalation enfilée* » ou « *intercalation en pas de vis* » se caractérise par des molécules possédant un substituant de chaque côté de leur partie plane aromatique qui normalement s’intercalera (Figure 2.5). En conséquence, afin de permettre l’intercalation du chromophore, l’un des substituants devra passer entre les bases empilées et atteindre l’autre côté de l’hélice. Les substituants qui devront passer entre les bases seront souvent volumineux et polaires, ce qui représentera une étape limitante dans la cinétique d’association des partenaires qui sera donc réduite de façon significative en comparaison des intercalants simples. Il en sera de même pour la dissociation. Le passage de gros substituants, impliquera également une ouverture plus importante des paires de bases qui pourront avoir lieu uniquement après d’importantes distorsions de la double hélice.

Parmi ce type d’intercalants, on pourra notamment citer le naphthalène-bisamide ou la nogalamycine. On retrouvera également certaines molécules poly-

intercalantes, pouvant intercaler trois [152], quatre [153], voir jusqu'à huit [154] plans aromatiques entre les paires bases de l'ADN.

2.2.3 Intérêts biochimiques et médicaux

Les acides nucléiques sont au cœur de très nombreux processus cellulaires. Un changement de leur structure induit par la présence de molécules intercalantes pourra de ce fait être à l'origine de nombreux dysfonctionnements à tous les niveaux.

Les acides nucléiques représenteront donc une cible naturelle pour de petites molécules d'intérêt thérapeutique dans le cadre de thérapies anticancéreuses ou anti-biotiques, mais également pour des molécules pouvant servir de photo-intercalants dans le cadre d'études expérimentales.

2.2.3.1 Interacalants et thérapie anticancer

Suite à l'élucidation de la structure de la double hélice et l'appréhension de son rôle clé au sein des êtres vivants, l'ADN est vite apparue comme une cible importante dans le traitement de maladies d'origines génétiques tels que les cancers. Le premier composé décrit comme ayant une action sur l'ADN fut le gaz moutarde [155]. Cependant, sa haute toxicité a poussé à la recherche de composés moins toxiques mais plus efficaces. Depuis la description du processus d'intercalation [146,156] et des premières molécules intercalantes, nombre de molécules de ce type se sont montrées efficaces dans les traitements de cancer.

Cependant, beaucoup d'intercalants classiques de l'ADN se sont avérés avoir de désastreux effets secondaires dus à leur manque de spécificité. De plus, pour avoir un réel intérêt dans les thérapies *in vivo* outre le fait de devoir présenter une certaine spécificité d'action, la molécule devra pouvoir outrepasser plusieurs barrières telles les membranes cytoplasmiques et nucléaires afin d'atteindre sa cible tout en restant stable en présence d'enzymes. La molécule devra ensuite, en plus

de s'intercaler, permettre d'inhiber certaines enzymes indispensables au bon fonctionnement des mécanismes essentiels à la vie cellulaire.

Parmi les molécules ayant une activité reconnue, de nombreuses cibleront les topoisomérases, ou leurs complexes avec l'ADN. On pourra notamment citer la famille des anthracyclines (voir section 2.2.3.3), l'étoposide, les acridines, les coumarines etc. (Voir Martinez et Chacon-Garcia 2005 [157] pour une revue).

Malgré le grand nombre de molécules disponibles à ce jour, de nombreuses présentent encore des effets secondaires non négligeables parmi lesquels on pourra trouver de simples nausées, des développements de leucémies, des effets cardiotoxiques, des lésions de moelles osseuses et des phénomènes de démyélinisation, etc [158,159]. De plus, les cellules développeront dans certains cas des résistances à ces agents chimiothérapeutiques telles des mutations de certains résidus de polymérase, ou encore des surexpressions de certaines protéines et transporteurs (notamment ceux impliqués dans le transport des protéines *multidrug* résistantes), etc [158,159].

C'est pourquoi, une compréhension approfondie des mécanismes d'actions de ces molécules intercalantes est aujourd'hui importante afin de développer de nouveaux agents intercalants plus efficaces.

2.2.3.2 Marquage

Certaines molécules intercalantes sont fluorescentes et donc utiles en tant que marqueur. A ce jour, il existe peu de molécules photointercalantes. On pourra notamment citer le DAPI utilisé comme marqueur en microscopie de fluorescence. Ces molécules présentent un grand intérêt notamment dans la compréhension des processus moléculaires guidant les intercalations [160].

2.2.3.3 Effets biologiques des anthracyclines – inhibition des topoisomérases

Les anthracyclines sont des molécules intercalantes qui, depuis leur découverte dans la première moitié des années 1960 [161,162], ont été intensivement utilisées en tant qu'agents thérapeutiques dans le traitement de grand nombre de cancers pour leur activité antinéoplasique (agit contre le développement de tissus tumoraux).

Dans la famille des anthracyclines, l'adriamycine et daunomycine (DNM) ont une structure identique à une fonction alcool prêt (l'adriamycine ayant un groupement OH en plus sur la chaîne latérale en C9). Cette différence a une grande influence sur le spectre d'activité de ces deux molécules. En effet, tandis que l'adriamycine sera effective dans les traitements du cancer du poumon, de lymphomes agressifs ou encore dans les cas de sarcomes des tissus mous, la DNM aura une application dans les cas de leucémie lymphoblastique ou myéloblastique.

Malgré une utilisation intensive de la famille des anthracyclines dans les traitements cliniques, leurs mécanismes d'actions et particulièrement ceux de la DNM et de l'adriamycine sont encore sujets à controverse [163,164].

Au-delà de ce débat qui ne concerne pas notre propos, la DNM au même titre que nombre d'autres dérivées de la même famille, est reconnue comme un « poison des topoisomérases II » [159,165]. Les topoisomérases, sont des enzymes contrôlant la structure topologique dont notamment le super-enroulement de l'ADN en générant des coupures transitoires d'un (topoisomérases de type I) ou des deux brins d'ADN (topoisomérases de type II) au niveau des liaisons phosphodiester. Ces molécules ont un rôle essentiel dans les étapes de réplication ou de transcription de l'ADN.

Les anthracyclines agiront en stabilisant un intermédiaire dans lequel l'ADN est coupé et lié de façon covalente au résidu tyrosine de la topoisomérase II, formant ainsi un complexe ternaire anthracycline-ADN-topoisomérase II inhibant ainsi la réplication cellulaire, conduisant à la mort de la cellule.

Parmi cette classe de médicament, la DNM a également été reconnue pour avoir une action sur les topoisomérases de type I [166]. Plus généralement les anthracyclines ont également des activités avérées d'activateur d'apoptose ou encore d'inhibiteur d'hélicases [159].

Au-delà de leur action et à de nombreux niveaux, les anthracyclines et principalement la daunomycine et l'adriamycine sont reconnues depuis de nombreuses années pour être des cardiotoxiques [159], et c'est pourquoi de nombreux travaux ont visé à en créer de nouveaux dérivés.

2.3 La daunomycine

La *daunomycine* (DNM) également connue sous le nom de *daunorubicine* fait partie, au même titre que l'*adriamycine* (ou également *doxorubicine*), de la famille des *anthracyclines*.

Les anthracyclines sont une large famille de composés d'origine bactérienne ayant une activité anticancéreuse et antibiotique dont les premiers dérivés furent isolés dans les années 1960 de façon quasi simultanée par deux équipes italienne et française, à partir d'actinobactéries productrices de pigment du genre *Streptomyces peucetius* [161,162].

Avec plus de 20 structures à haute résolution ainsi que de nombreuses études cinétiques et thermodynamiques concernant son intercalation, la DNM fait partie des molécules intercalantes ayant le plus de données disponibles à ce jour [167]. En effet, de nombreux détails concernant le mode et la spécificité d'interaction de la DNM avec l'ADN ont été obtenus par des techniques de cristallographie aux rayons X [168,169] ainsi que par spectroscopie RMN [170,171]. Des études physico-chimiques, basées sur des techniques d'absorbance [172], de dialyse [172], de fluorescence [172-175], de dichroïsme circulaire [176] ou encore de footprinting [177], ont permis d'avoir des approches quantitatives de la relation structure activité de la DNM en solution aqueuse.

La DNM au même titre que de nombreuses molécules intercalantes, s'intercalera de façon préférentielle entre des paires de bases (A/T)pCpG.

2.3.1 Structure libre et intercalée

La structure de la DNM libre ou intercalée dans l'ADN a été déterminée par des techniques de cristallographie aux rayons X. Les premières structures de la DNM libre furent résolues à la fin des années 1970 [178-180]. Il fallut ensuite attendre une dizaine d'années et les travaux de Wang et Rich [181] en 1987, pour voir apparaître la première structure cristallographique d'un complexe entre la DNM et un hexamère d'ADN. Ce fut d'ailleurs la première structure résolue d'un mono-intercalant interagissant avec l'ADN. De nombreuses structures de la DNM complexée à l'ADN furent publiées par la suite [168,169,182-188].

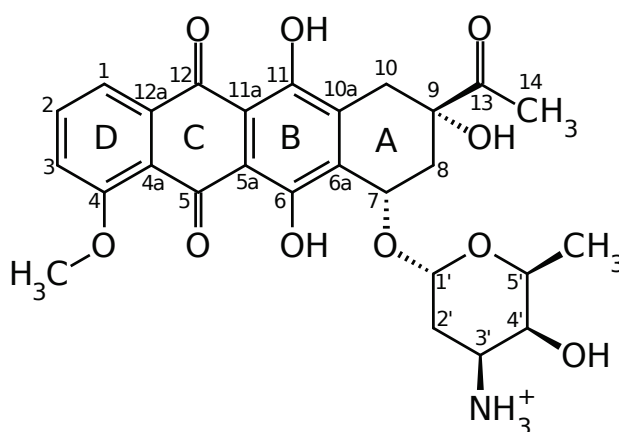


Figure 2.6– Représentation schématique de la Structure chimique de la daunomycine.

La DNM, dont la structure est représentée par la Figure 2.6, est constituée d'un chromophore aglycone composé de quatre cycles accolés (un cycle A flexible et des

cycles B, C et D formant une anthraquinone). Les cycles B, C et D correspondent à la partie plane de la molécule qui s'intercale. Sur le C7 du cycle A est lié un sucre aminé de type daunosamine. Dans les complexes DNM-ADN, ce sucre se trouve dans le petit sillon de la double hélice [181]. Le cycle A porte également un groupe acétyle et un groupement alcool, tous deux liés au C9. Un groupement méthoxy lié au cycle D (lié en C4) pointe dans le grand sillon lors de la formation d'un complexe DNM-ADN.

Le passage de la DNM d'une forme libre dans le solvant à une forme complexée avec l'ADN, engendrera des déformations structurelles des deux partenaires. Ainsi, on pourra noter des changements de conformation au niveau du cycle A et du sucre aminé de la DNM afin de faciliter l'insertion dans la cavité serrée que représente le site d'intercalation.

De même, l'insertion de la DNM influencera la structure de la double hélice. La formation de la cavité d'intercalation nécessitera une augmentation de la séparation des paires de bases au niveau du site d'intercalation (qui sera principalement entre des pas CpG pour la DNM), ce qui entraînera naturellement une déformation du paramètre de *rise* de l'ADN. Dans le cadre de leur étude du chemin d'intercalation, Mukherjee *et al.* [227], observent une augmentation de la valeur de *rise* de plus de 4 Å par rapport à un ADN-B classique (dont le *rise* est de ~3.4 Å), lorsque la DNM est intercalée.

D'après les études expérimentales, l'intercalation de molécules dans l'ADN aura également une influence sur le paramètre de *twist* de l'ADN, au niveau du pas d'intercalation lui-même ou sur les pas adjacents. En comparaison avec un ADN-B canonique (dont la valeur de *twist* sera d'environ 34°), l'intercalation engendrera une diminution du *twist* dont la magnitude dépendra de molécule intercalée, comme nous l'avons déjà remarqué [181,189,190]. Dans le cas de la DNM, les structures cristallographiques montrent une réduction du *twist* global de 8° (pour une séquence de 6 paires de bases), localisée au niveau des deux paires de bases adjacentes au site d'intercalation. Les paires de bases impliquées dans l'intercalation présentent cependant des valeurs de *twist* normales [168,181].

D'autres études basées sur le coefficient de sédimentation de l'ADN en présence de divers intercalants [191] aboutissent à une diminution du *twist* d'environ 11° en présence de DNM, mais sans pouvoir localiser précisément les bases concernées par cette déformation. On pourra également citer des travaux de DM dont notamment ceux de Trieb *et al.*, en 2004 [192]. Leurs simulations, ont permis d'observer que l'intercalation d'une seule DNM était à l'origine d'une forte diminution du *twist* (de l'ordre de 26° de moins par rapport à un ADN-B) au niveau du site d'intercalation, tandis que dans le cas d'une double intercalation cette diminution est moins importante (entre 12° et 14°). En 2008, lors de l'étude du chemin d'intercalation de la DNM, Mukherjee *et al.*, ont quant à eux observé une valeur de *twist* de l'ordre de 17° au niveau des paires de bases impliquées dans l'intercalation.

2.3.2 Flexibilité du cycle A

Le cycle A de la DNM présente une flexibilité au niveau de l'angle de torsion C7-C8-C9-C10 comparable aux carbohydrates cycliques. Les conformations du cycle A de la DNM ainsi que de ces analogues ont été étudiées par des techniques de cristallographie [168,169,178,179,182-188] ainsi que par des techniques de spectroscopie RMN du proton en une ou deux dimensions et dans différents solvants [171,193-200]. Il est ressorti de ces études qu'il pouvait exister différents conformères et que les positions des atomes C8 et C9 pouvaient servir pour les classer selon la nomenclature généralement utilisée pour les carbohydrates [201], illustrée par la Figure 2.7.

Les analyses cristallographiques de la DNM libre ont permis de montrer que le cycle A se trouvait dans une conformation ⁹H₈ [178,179]. Dans le cas de la DNM complexée avec l'ADN, Nunn a observé différentes conformations en demi-chaise du cycle A semblant être corrélées avec le type de base entourant la DNM. Ces conformères étant de types ⁹H₈, ⁹S et S₈ [169]. Les mêmes types de conformations sont observés dans les autres études, mais sans notion de dépendance sur la séquence

de bases. Dans toutes les études cristallographiques, le cycle A de la DNM présente une conformation avec le groupement acétyle placé en position équatoriale correspondant à une conformation 9H_8 (Figure 2.7).

Les premières analyses par spectroscopie RMN arrivèrent à la conclusion que dans les deux cas (libre ou complexé) le groupement acétyle du cycle A de la DNM, contrairement aux analyses cristallographiques, se trouvait en position axiale [193,194]. Ces résultats furent appuyés par des analyses de résonance paramagnétique [202].

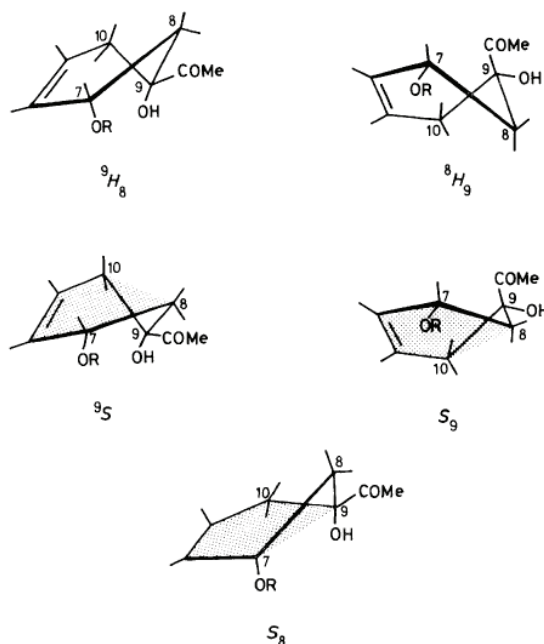


Figure 2.7 – Conformations potentielles du cycle A de la DNM. Dans les conformations 9H_8 , 9S (ou 9H) et S_8 (ou H_8) le groupement acétyle en C9 sera en position équatoriale tandis qu'avec les conformations 8H_9 et S_9 (ou H_9), il sera en position axiale. (Issu de [196]).

Par la suite, en 1987, Mondelli *et al.* [195,196], à travers leurs études des conformations du cycle A de la DNM (et de plusieurs analogues) en présence de différents solvants, arrivèrent à la conclusion que quelque soit le solvant utilisé, le cycle A adoptait préférentiellement une conformation en demi-chaise 8H_9 . Par la

suite, une série d'études de RMN 1D et 2D menée par Barthwal [171,197-200] et ses collaborateurs est venue appuyer les résultats de Mondelli. Dans leur dernière étude datant de 2008, Barthwal *et al.* [200], ont également démontré que leurs résultats obtenus par RMN étaient corrélés avec des approches de mécanique quantique.

Toutes ces études ont donc permis de conclure, que libre ou complexé, le cycle A de la DNM adopte préférentiellement une conformation en demi-chaise 8H_9 , avec l'angle dièdre C7-C8-C9-C10 prenant une valeur d'environ 55°.

2.4 Aspects thermodynamiques et cinétiques

Malgré le nombre d'études concernant les complexes d'intercalations, la nature des facteurs énergétiques guidant ce processus n'est pas encore appréhendée dans son ensemble [160]. Nombre d'« évènements » moléculaires contribuent à l'énergie libre (ΔG) de la réaction d'intercalation, certains d'entre eux étant enthalpiques (ΔH) et d'autres entropiques (ΔS). Ils sont reliés à l'énergie libre selon la relation :

$$\Delta G = \Delta H - T\Delta S$$

Les premières tentatives pour obtenir l'énergie libre d'intercalation [203] ont permis de révéler que ce type de processus était accompagné par un changement négatif de la capacité calorifique suggérant un rôle important des interactions hydrophobes guidant la formation de tels complexes [204].

Les intercalants se lient à l'ADN en s'insérant entre deux paires de bases. Leur interaction avec l'ADN nécessite de ce fait, la formation d'une cavité d'intercalation s'accompagnant de changements conformationnels de l'ADN (déroulement et élongation de la double hélice) [205,206]. Une fois inséré dans la cavité, l'intercalant pourra former des interactions supplémentaires (liaisons hydrogène, van der Waals, etc.) pour fixer le complexe.

Une étude des données thermodynamiques d'une dizaine de molécules intercalantes publiée par Chaires en 2006 [207], a notamment permis de montrer que le processus d'intercalation est guidé par une large contribution enthalpique. En effet, tous les intercalants de cette étude présentent des valeurs négatives d'enthalpie, tandis qu'ils présentent des contributions entropiques défavorables (c'est-à-dire nulles ou positives).

Cette entropie défavorable est le résultat de plusieurs contributions. La formation de complexes biomoléculaires sera à l'origine d'une réduction des degrés de liberté translationnels et rotationnels impliquant une pénalité entropique dont la magnitude est controversée [204,208-211]. Les molécules intercalantes, à l'instar de nombreuses molécules interagissant avec l'ADN, sont souvent positivement chargées. L'interaction de telles molécules avec l'ADN résultera en un relargage d'un contre-ion, ce qui, en accord avec la théorie des polyélectrolytes de Manning et Record [107,212] sera à l'origine d'une contribution entropique favorable dont la magnitude sera dépendante de la charge du ligand [213]. Une contribution positive supplémentaire sera apportée par la formation de la cavité nécessaire à l'intercalation elle-même. En effet, la création de cette cavité, de par les déformations de la double hélice qu'elle engendrera, augmentera la rigidité de l'ADN résultant en une entropie défavorable [207].

Concernant la composante enthalpique, les études des complexes protéines-ADN ont permis de montrer que les interactions impliquant une déformation de l'ADN étaient souvent associées à une enthalpie défavorable [214]. A contrario, lors de l'intercalation, malgré la déformation de la double hélice au niveau du site d'intercalation, la composante enthalpique apporte une contribution favorable à l'énergie libre. En effet, la formation de la cavité d'intercalation à l'origine de la rupture des interactions d'empilements entre les paires de bases enthalpiquement défavorables, sera compensée lors de l'intercalation, par la formation de nouvelles interactions de type π -stacking entre les bases et le chromophore aromatique. Toutes les autres interactions non-covalentes permettant de fixer l'intercalant dans son site d'intercalation (telles les effets hydrophobes, des interactions élec-

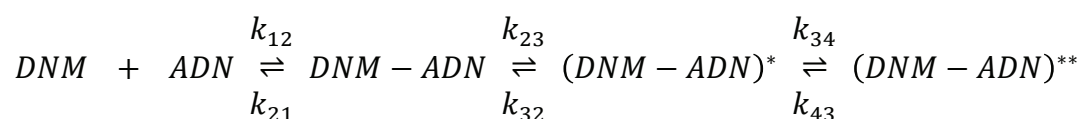
trostatiques, de van der Waals, et les liaisons hydrogène [146,156,215,216]) contribueront également à une contribution enthalpique favorable à l'énergie libre d'intercalation.

Il existe de nombreuses études thermodynamiques concernant la DNM et ses dérivées donnant la variation d'énergie libre entre leur état libre et leur état intercalé. On pourra notamment citer les travaux de Chaires et ses collaborateurs de 1996 dans lesquels ils étudièrent la liaison de la DNM à l'ADN en fonction de la concentration en sels [217] ; mais également les travaux de Remeta *et al.*, [218] qui ont permis de décrire des énergies libres de formations de complexes variant selon la séquence entre -10,8 et -7,5 kcal mol⁻¹ ; ou encore l'étude de Ren *et al.* en 2000 [203].

Ces études thermodynamiques seront incapables par elles-mêmes de définir les facteurs dirigeant les processus de reconnaissance et d'intercalation, et il sera donc nécessaire de les compléter par des études cinétiques et structurales.

Les études portant sur la cinétique de l'intercalation sont, quant à elles, moins courantes, et pour l'instant le mécanisme guidant l'intercalation n'est pas encore clairement défini. En effet, deux modèles sont généralement admis :

Les travaux menés par Chaires et ses collaborateurs, basés sur des techniques d'absorption, décrivent l'intercalation comme un processus s'effectuant en trois étapes [219] selon un schéma :



Ce schéma pourra être décrit de la façon suivante : une première étape « bimoléculaire » de formation de liaisons faibles à « l'extérieur » de l'ADN permettant de créer le complexe *DNM-ADN* non intercalé. Cet état pouvant potentiellement permettre des changements conformationnels dont la formation de la cavité d'intercalation en vue de l'insertion de la molécule. Cette première étape possède une constante de vitesse de second ordre de 3.10⁶ M⁻¹ s⁻¹. Puis une seconde étape « unimoléculaire » d'intercalation de la molécule dans l'ADN représentée par

l'intermédiaire (*DNM-ADN*)*. S'ensuivra une troisième étape de réarrangement de la DNM ou de l'ADN ainsi que la formation d'interactions moléculaires supplémentaires (tels des liaisons hydrogène, d'empilements de vdW, etc. [207]) représentée par le complexe (*DNM-ADN*)** sur le schéma proposé.

En comparaison, Rizzo *et al.*, suite à leurs travaux, proposent quant à eux un modèle cinétique en cinq étapes [220]. Ce modèle diffère de celui de Chaires par le fait qu'il propose deux états bi-moléculaires : un état où les partenaires sont faiblement liés, qui ne conduira pas directement à l'état intercalé ; un état conduisant à l'intercalation et présentant une constante de vitesse de second ordre de $6.10^6 \text{ M}^{-1} \text{ s}^{-1}$. Les autres étapes du modèle étant associées à des relaxations de la DNM ou de l'ADN.

Un petit nombre de travaux ont également tenté d'étudier le mécanisme d'interaction par une approche théorique [215,221] tout en se concentrant principalement sur le calcul des variations d'énergie libre totale en utilisant une représentation implicite de l'environnement [222-226], et ce grâce à des méthodes simplifiées de calcul des contributions électrostatiques et non-électrostatiques de l'énergie libre. Cependant, en plus d'ignorer le rôle de l'eau, ces approches négligeront de larges contributions (telles les contributions électrostatiques et non-électrostatiques à l'énergie libre, les pertes d'entropie, etc.) aboutissant à une variation d'énergie libre totale plus petite.

Une étude récente de dynamique moléculaire en solvant explicite, menée par Mukherjee et ses collaborateurs [227], a permis d'apporter quelques lumières sur le processus d'intercalation de la DNM, ainsi que sur l'énergie libre le long d'un chemin allant de la DNM libre dans le solvant à une position intercalée, en passant par un état lié au petit sillon. Les énergies libres calculées sont proches de celles obtenues par les méthodes expérimentales.

CHAPITRE 3

Méthodes

3.1 D'une vision rigide à une vision dynamique : introduction à la dynamique moléculaire

De nos jours, grâce à des techniques expérimentales de résolution de structure telles la cristallographie aux rayons X ou la résonance magnétique nucléaire (RMN), il est possible d'obtenir de nombreuses structures moléculaires (médicaments, protéines, acides nucléiques, complexes protéiques, etc.). Résoudre une structure consiste à déterminer les coordonnées tridimensionnelles des atomes de la molécule d'intérêt (x, y, z). Cependant, la résolution de structure fournit une vision statique des molécules. Cette vision statique doit être remplacée par une représentation dynamique des systèmes moléculaires avec les atomes en mouvements constants. Ces mouvements atomiques peuvent être locaux et rapides, ou collectifs et plus lents, impliquant des groupes d'atomes, voire des domaines entiers de la molécule, le tout se résumant à un ensemble de conformations possibles pour une molécule donnée à une température donnée.

Types de mouvements	Etendue spatiale (nm)	Amplitude (nm)	Temps caractéristiques (s)
Vibration relative d'atomes liés	0.2-0.5	0.001-0.01	10^{-14} - 10^{-13}
Vibration élastique de région globulaire	1-2	0.005-0.05	10^{-12} - 10^{-11}
Rotation des chaînes latérales en surface	0.5-1	0.5-1	10^{-11} - 10^{-10}
Vibration de torsion de groupes enfouis	0.5-1	0.05	10^{-11} - 10^{-9}
Mouvement relatif de différentes régions globulaires (charnière et courbure)	1-2	0.1-0.5	10^{-11} - 10^{-7}
Rotation de chaînes latérales de taille moyenne à l'intérieur de la protéine	0.5	0.5	10^{-4} -1
Transitions allostériques	0.5-4	0.1-0.5	10^{-5} -1
Dénaturation locale	0.5-1	0.5-1	10^{-5} -10

Tableau 3.1 – Caractéristiques de quelques mouvements de protéines (adaptés de McCammon, J. A. Protein dynamics. 1984 [228]).

Différentes techniques expérimentales (telles la spectroscopie de fluorescence, la RMN, etc.) permettent de sonder la dynamique des protéines sur des échelles de temps pouvant aller de la femtoseconde à la microseconde. Cependant, ces méthodes expérimentales ne donneront qu'une vision très partielle de la dynamique des systèmes moléculaires, pouvant eux couvrir des échelles de temps allant de la femtoseconde à la seconde, voire plusieurs dizaines de secondes (voir Tableau 3.1). Des méthodes alternatives aux études *in vitro* des mouvements des protéines consistent à simuler *in silico* ces mouvements par le biais de la modélisation moléculaire.

La modélisation, en complément des techniques expérimentales, permet une compréhension et, dans les meilleurs cas, une prédiction du comportement de la matière dans différentes conditions définies (températures ou pressions données par exemple). Dans ce sens, elle s'appuiera sur des méthodes de calculs théoriques dont les principales vont faire appel à la mécanique quantique, la mécanique moléculaire ainsi qu'à la dynamique moléculaire afin de déterminer les propriétés physico-chimiques et géométriques des atomes d'un système [229].

La Dynamique Moléculaire (DM) permet, à partir des lois de la mécanique classique, de simuler la trajectoire des atomes d'un système moléculaire en solution, en phase cristalline ou en phase gazeuse [229-231], et ainsi de fournir des informations sur l'évolution du système au cours du temps. L'énergie du système est déterminée via une résolution numérique des équations classiques du mouvement.

Une molécule est considérée comme un ensemble d'atomes avec des interactions intermoléculaires et intramoléculaires. L'énergie potentielle de la molécule est une somme de termes dépendant de plusieurs paramètres, tels que la position et la nature des atomes. La forme analytique de l'énergie potentielle (la fonction d'énergie potentielle) et le jeu de paramètres d'interactions interatomiques nécessaires à son calcul constituent le « champ de forces ». Dans un champ de forces classique, les atomes seront représentés par des masses ponctuelles chargées et reliées par des ressorts (dans le cas des liaisons covalentes), le mouvement des électrons étant quant à lui ignoré.

3.2 Champs de forces

Apparus dans les années 1970, les champs de forces sont en constante évolution afin d'obtenir des simulations numériques les plus proches possibles de la réalité.

3.2.1 Expression du champ de forces

Le champ de forces AMBER utilise une représentation empirique de l'énergie dont l'expression est la suivante :

$$E_{total} = \sum_{liaisons} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dièdres} \frac{V_n}{2} [1 + \cos(n\varphi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] \quad (3.1)$$

Cette description du potentiel provient de Cornell *et al*, [232]. Cette énergie se décompose en composantes intramoléculaires et intermoléculaires.

3.2.2 Interactions intramoléculaires – « liées »

Les deux premiers termes du champ de force correspondent aux énergies potentielles de déformation des longueurs et des angles des liaisons. Ces deux énergies sont représentées par un potentiel harmonique avec une sommation sur toutes les liaisons ou sur tous les angles entre atomes liés par liaisons covalentes.

Terme de déformation des liaisons

$$\sum_{\text{liaisons}} k_b (b - b_0)^2 \quad (3.2)$$

Dans cette expression, k_b correspond à la constante de forces associée à cette liaison (en kcal mol⁻¹.Å⁻²), b à la longueur de liaison instantanée (en Å) et b_0 à la longueur de liaison à l'équilibre.

Terme de déformation des angles

$$\sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 \quad (3.3)$$

Ici, k_θ correspond à la constante de forces associée à cet angle (en kcal mol⁻¹), θ à l'angle instantané (en °) et θ_0 à la valeur à l'équilibre.

Le troisième terme décrit l'énergie de déformation des angles dièdres, son expression étant définie par un développement en série de Fourier.

Terme de déformation des dièdres

$$\sum_{\text{dièdres}} \frac{V_n}{2} [1 + \cos(n\varphi - \gamma)] \quad (3.4)$$

φ correspondant à la valeur instantanée de l'angle dièdre (en °), n représentant la multiplicité, γ la phase (en °) et V_n correspondant à la barrière de potentiel (en kcal mol⁻¹).

Ce terme faisant intervenir quatre atomes, il prend déjà en compte une partie des interactions entre atomes non liés qui devront donc être pondérés. On ajoute également à ce terme une partie correspondant aux angles impropres afin d'assurer la planéité de certains groupements conjugués.

3.2.3 Interactions intermoléculaires – « *non liées* »

La dernière partie de l'expression (3.1) correspond à la contribution de l'énergie des interactions *non liées*. Ces interactions *non liées* concernent les atomes qui appartiennent à des molécules différentes ou qui sont séparés par plus de trois liaisons covalentes. Les interactions sont décrites par deux termes : un terme de type Lennard-Jones (pour les interactions de type van der Waals (vdW)), et un terme électrostatique.

3.2.3.1 Interactions de Van der Waals

Le terme de Lennard-Jones est représenté par la somme des énergies résultantes des interactions attractives et répulsives. Les interactions de vdW sont représentées par un potentiel en 6-12 dont l'expression est la suivante :

$$\sum_{i < j} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \quad (3.5)$$

Dans cette expression, A_{ij} correspond aux paramètres des termes répulsifs à courte distance (reflets de la répulsion des nuages électroniques) et B_{ij} est le paramètre des termes attractifs à une distance moyenne (forces de dispersion de London, résultant de l'induction et de l'attraction de dipôles instantanés).

3.2.3.2 Interactions électrostatiques

Le terme électrostatique du champ de forces est de la forme :

$$\sum_{i < j} \frac{q_i q_j}{\epsilon R_{ij}} \quad (3.6)$$

L'énergie électrostatique est représentée par un potentiel coulombien où ϵ est la constante diélectrique, q_i et q_j sont les charges partielles des atomes i et j et R_{ij} leur distance.

La densité électronique d'une molécule est décrite à l'aide de charges atomiques, déterminées par des calculs sur de petites molécules via l'utilisation de méthodes quantiques semi-empiriques ou *ab initio*. Dans le champ de forces que nous utilisons pour nos calculs qui est parm99, la méthode de répartition des charges est basée sur la procédure RESP (pour *Restrained Electrostatic Potential*) [233] qui consiste à dériver les charges du potentiel électrostatique *ab initio* au niveau Hartree-Fock en base 6-31G*.

3.2.4 De la formulation à l'utilisation pour l'ADN

L'ADN et les acides nucléiques en général, du fait de leur grand nombre de charges ainsi que de leur grande flexibilité ont pendant longtemps représenté un véritable défi pour leur intégration et leur juste description dans les champs de forces. En effet, bien que la première simulation de dynamique moléculaire de protéine fut publiée dans les années 1970 [234], il fallut attendre le milieu des années 80 pour voir la première simulation contrainte d'acides nucléiques [235,236]. Il fallut attendre encore dix ans et l'avènement des nouveaux champs de forces ainsi que l'incorporation de nouvelles méthodes de représentation des effets électrostatiques à longue distance pour voir la première simulation d'acides nucléiques sans contraintes [76,232,237-240].

Bien que d'autres soit disponibles, les champs de forces implémentés dans les logiciels de dynamique moléculaire CHARMM (*Chemistry at HARvard Molecular Mechanics*) et AMBER (*Assisted Model Building with Energy Refinement*) sont ceux qui ont été les plus utilisés [76,232,240,241]. Les simulations de dynamique moléculaire ont montré une bonne reproductibilité de propriétés structurales et dynamiques des acides nucléiques en solution [68,117,242-244], notamment avec les dernières versions qui sont habituellement utilisées pour les acides nucléiques :

parm94 et parm99, ces champs de forces montrant un bon accord structurel de l'ADN entre les structures calculées par dynamique moléculaire et celles obtenues dans les cristaux ou en solution [245].

Basée sur des observations de longues simulations de dynamique moléculaire d'ADN-B montrant des surpopulations de conformations non canoniques des angles α et γ [72,73,75,137], une correction récente du champ de forces parm99 nommée bsc0 a été ajoutée [75]. Cette modification qui a consisté à reparamétriser parm99, a permis d'affiner la description de l'espace conformationnel des angles α et γ et donc d'améliorer la stabilité des ADN-B canoniques lors de longues simulations.

Au vu de ces différents arguments, nous avons donc décidé d'effectuer nos calculs de dynamique moléculaire avec le champ de forces parm99 avec la modification bsc0.

3.3 Traitement du solvant et des ions

3.3.1 Quelle représentation de l'environnement ?

Etant un polyélectrolyte fortement chargé, l'ADN est particulièrement sensible à l'environnement qui l'entoure. C'est pourquoi une représentation adaptée de l'environnement est nécessaire pour caractériser ses propriétés lors des simulations. Néanmoins, compte tenu des ressources informatiques disponibles à l'heure actuelle il n'est pas encore possible de considérer entièrement l'environnement physiologique, composé non seulement de l'eau et de sels, mais aussi de multiples macromolécules et assemblages macromoléculaires.

3.3.1.1 Représentation implicite

Les systèmes macromoléculaires impliquant les acides nucléiques existent dans un environnement aqueux. Il est donc justifié de considérer que de tels systèmes seront représentés solvatés dans de l'eau pendant les simulations. Cependant, une grande partie des temps de calculs est utilisée pour évaluer les interactions des molécules de solvant entre elles. Il est donc apparu utile d'éviter de représenter le solvant de façon explicite tout en se rappelant que les effets de solvant sont essentiels. C'est pourquoi de nombreux modèles de solvant implicites ont été développés [246-249].

Les représentations implicites vont permettre une bonne estimation de l'énergie libre de solvation [250] ainsi que rendre possible une description bien plus rapide de grands systèmes biologiques qui pour l'instant sont encore trop coûteux voire inaccessibles par des traitements explicites du solvant en dynamique moléculaire.

De plus, de nombreuses avancées ont été réalisées dans les représentations implicites, notamment dans la description des solvants via l'exploitation des modèles types Poisson-Boltzmann (PB) ou Generalized-Born (GB) [248] (cette dernière méthode permettant d'ailleurs d'effectuer des DM bien plus rapidement qu'avec une représentation explicite du solvant [251]) ainsi que le développement de logiciels associant solvant implicite et dynamique stochastique, tels SDA [252], UHBD [253], BrownDye [254], etc.

3.3.1.2 Représentation explicite

Le solvant qui entoure les macromolécules biologiques aura une grande influence sur la structure, les propriétés physico-chimiques et les interactions de ces systèmes. C'est pourquoi, si l'on veut comprendre ces systèmes, il faut les considérer dans leur ensemble. Représenter le solvant de façon explicite paraît donc le moyen le plus adapté d'appréhender toutes les propriétés de ces systèmes. Le

temps de calcul étant proportionnel à $N \cdot \log N$ (N représentant le nombre d'atome du système), pendant longtemps la représentation explicite du solvant a représenté un frein à l'étude de gros systèmes par dynamique moléculaire, les simulations étant limitées par le grand nombre de molécules d'eau nécessaires pour hydrater de gros systèmes. De nos jours, les progrès des techniques informatiques et notamment la grande augmentation des puissances de calculs (on pourra notamment citer le supercalculateur Anton [255]) ont rendu possible l'étude de gros systèmes avec des représentations explicites du solvant, introduisant à la fois un grand nombre de molécules d'eau et des ions (tels Na^+ , K^+ , Cl^- ou encore Mg^{2+}) et permettent ainsi une étude dans des conditions plus proches des conditions physiologiques.

Comme nous l'avons vu précédemment, les couches d'hydratation ont un rôle important dans la structure des acides nucléiques (voir partie 1.4.1). Malgré de nombreux efforts de développement [256,257] l'utilisation de solvants implicites est encore peu recommandée pour décrire les systèmes impliquant des acides nucléiques [258]. De ce fait, dans tous nos calculs de dynamique moléculaire, le solvant sera représenté de façon explicite.

3.3.2 Modèles de représentation de l'eau

Depuis les origines de la dynamique moléculaire, de nombreux modèles d'eau ont été proposés. L'efficacité avec laquelle l'énergie peut être calculée avec un modèle donné est souvent un facteur important puisque dans une simulation de nombreuses molécules d'eau sont présentes autour du soluté à simuler. Les modèles d'eau peuvent être divisés en trois types : les modèles de simple interaction rigide, les modèles flexibles ainsi que les modèles incluant les effets de polarisation.

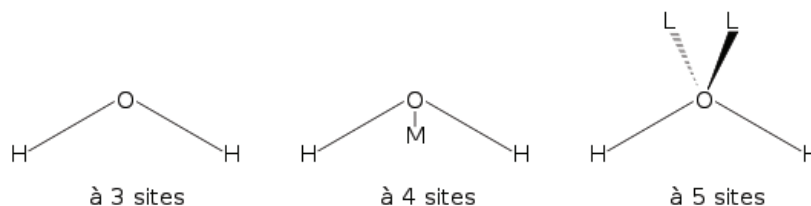


Figure 3.1 – Modèles de représentation des molécules d’eau. Modèle à 3 sites : les charges partielles positives sont localisées sur les hydrogènes, et la charge négative est portée par l’oxygène. Modèle à 4 sites : la charge négative de l’oxygène est déplacée sur un pseudoatome noté M se trouvant sur la bissectrice de l’angle H-O-H. Modèle à 5 sites : la charge négative est placée sur deux pseudoatomes (notés L) reliés à l’oxygène reproduisant ainsi le doublet non liant de l’atome d’oxygène.

Dans le cas du modèle dit de « simples » interactions, chaque molécule d’eau est maintenue dans une géométrie rigide. Les interactions non liantes sont décrites par des termes coulombiens pour les interactions électrostatiques, et un terme Lennard-Jones pour les interactions de vdW. Ce modèle simple utilise entre trois et cinq sites d’interaction. Parmi les plus utilisés, on pourra citer les modèles TIP3P [259,260] et SPC [261] utilisant trois sites pour les interactions électrostatiques (Figure 3.1 à gauche). Les interactions de vdW sont centrées sur l’oxygène. Ces modèles différeront essentiellement d’un point de vue géométrique, au niveau des charges des hydrogènes ou dans les paramètres de Lennard-Jones. Dans les modèles simples on pourra également citer ceux à quatre sites d’interaction tel le modèle TIP4P [260] (voir Figure 3.1 au centre) qui permettra d’améliorer la distribution électrostatique autour de la molécule d’eau, ou encore les modèles à cinq sites d’interaction comprenant notamment les modèles ST2 [262] ou TIP5P [263] (voir Figure 3.1 à droite) améliorant la géométrie des dimères d’eau, mais plus coûteux en temps de calcul.

Une représentation rigide de l’eau reste une approximation, ce qui impliquera la perte de nombreuses propriétés. D’autres modèles dont notamment des modèles flexibles permettant les vibrations des molécules d’eau ou encore des mo-

dèles incluant de façon explicite les effets de polarisation ont été développés. Ces modèles ne seront pas développés dans notre propos.

Les champs de forces d'AMBER ont été paramétrés afin d'utiliser une représentation explicite des solvants. Dans le cas de parm99, les modèles d'eau les plus adaptés sont les modèles TIP3P, TIP4P et SPC/E [264]. Dans le cas du modèle TIP3P, chaque molécule sera représentée par trois points (deux hydrogènes et un oxygène) dont les atomes seront maintenus rigides par trois liaisons à des distances de 0.957 Å et 1.514 Å pour respectivement deux liaisons O-H une pseudo-liaison H-H par une constante de force de 553 kcal mol⁻¹ Å⁻².

Dans toutes les simulations de dynamique moléculaire, nous avons choisi le modèle TIP3P pour représenter explicitement l'eau dans le champ de forces AMBER parm99.

3.3.3 Les ions

In vivo, chaque molécule se trouve dans un milieu composé de nombreux ions. Il ne sera donc pas possible de proposer des simulations reproduisant de façon réaliste le comportement des molécules en milieu cellulaire sans notamment une bonne représentation des ions présents dans le milieu, ainsi que de leurs effets sur les solutés.

Lors du développement des champs de forces, de nombreuses attentions ont donc été portées à la description des ions notamment concernant les paramètres de Lennard-Jones des ions monovalents tels K⁺, Na⁺ ou Cl⁻ [265-267].

Des travaux ont démontré que les paramètres d'ions actuellement implémentés dans les champs de forces AMBER (paramètres tirés de Aqvist [265]) entraînaient une agrégation des ions à proximité des macromolécules étudiées dans des conditions de fortes concentrations (aux alentours de 1 M pour NaCl et KCl). Ces valeurs limites restant largement inférieures aux valeurs limite expérimentales

d'agrégation (3.2 M pour NaCl [268] et 5.4 M pour KCl [268]) et ce indépendamment du modèle d'eau testé (TIP3P ou SPC).

Une étude récente de Auffinger *et al.* [269], a permis de montrer que la meilleure alternative actuelle afin d'éviter ce type d'artéfacts était l'utilisation de paramètres Lennard-Jones définis par Dang en 1995 [270] (nous ferons référence à ces paramètres en tant que « paramètres de Dang ») et que de plus, l'utilisation des paramètres implémentés par défaut dans parm99 devait être évitée.

Généralement, les études de dynamique moléculaire impliquant les acides nucléiques sont menées dans des conditions de concentration en ions minimale, c'est-à-dire avec une concentration d'ions neutralisant exactement le milieu, ces conditions de simulation étant considérées comme satisfaisantes. Pourtant, ces conditions ne sont pas en adéquation avec le milieu physiologique dans lequel évoluent les acides nucléiques (environ 0.150 M de sel monovalent).

De ce fait, nos simulations ont été effectuées avec des ions paramétrés avec les paramètres de Dang dans des conditions de concentrations salines minimales ainsi que "physiologiques", ce qui nous permettra d'évaluer une éventuelle influence de la concentration ionique sur nos systèmes.

3.4 Simulations de dynamique moléculaire

La mise en œuvre d'un processus de dynamique moléculaire va nécessiter plusieurs étapes. Après une étape de construction du système initial consistant à placer la molécule à étudier dans une boîte de solvant, l'énergie potentielle de l'ensemble est ensuite minimisée afin de relaxer les tensions initiales avant de commencer la dynamique moléculaire à proprement parler.

3.4.1 Minimisation de l'énergie

Afin d'éviter les instabilités liées à la nécessité de relaxer une trop forte énergie potentielle initiale, il est nécessaire de minimiser la structure initiale. Cette étape va avoir pour but d'éliminer les mauvais contacts interatomiques et d'optimiser la structure de départ. Les structures ainsi minimisées correspondront au minimum local le plus proche de la structure de départ.

Différentes techniques de minimisation existent. Classiquement, deux méthodes sont utilisées : la méthode dite « *steepest descent* » ou « *méthode de la plus grande pente* » et la méthode dite « *conjugate gradients* » ou méthode des « *gradients conjugués* ». Ces techniques sont basées sur le calcul de dérivées de l'énergie potentielle U par rapport aux coordonnées de chaque atome i et des processus itératifs.

$$\vec{F}_i = -\overrightarrow{\text{grad}}U_i \quad (3.7)$$

Chaque atome sera déplacé selon la force qui s'exerce sur lui suivant la pente de la surface d'énergie, atteignant ainsi le minimum local le plus proche.

A chaque itération (ou pas de minimisation) k , on détermine le système de N atomes décrits par un vecteur X_k , la valeur du déplacement λ_k ainsi que la direction du déplacement S_k . Le nouveau jeu coordonné des atomes X_{k+1} au pas suivant pourra être défini par la relation suivante :

$$\overrightarrow{X_{k+1}} = \overrightarrow{X_k} + \lambda_k \overrightarrow{S_{k+1}} \quad (3.8)$$

Les deux méthodes citées précédemment vont différer dans le choix du vecteur direction ainsi que dans leur vitesse de convergence.

La méthode de la plus grande pente utilise une direction de recherche S_k dans le sens du gradient négatif (c'est à dire dans la direction de la force).

$$\text{Soit dans la direction : } \overrightarrow{S_k} = -\overrightarrow{g_k} \quad (3.9)$$

La recherche se fait donc en suivant la direction où la fonction d'énergie décroît le plus. Le pas sera ajusté pour que la recherche soit rapide, mais toujours dans la direction de l'énergie décroissante. Cette méthode sera donc très efficace pour s'approcher très rapidement d'un minimum local, cependant elle aura une convergence très lente aux environs des minima locaux. Elle sera le plus souvent utilisée comme première étape dans les processus de minimisation.

La méthode des gradients conjugués quant à elle, convergera rapidement à proximité du minimum. Dans ce cas, le gradient calculé à chaque pas sera conjugué au précédent par la relation suivante :

$$\vec{S}_k = -\vec{g}_k + b_k \vec{S}_{k-1} \quad (3.10)$$

avec

$$b_k = \frac{\vec{g}_k \cdot \vec{g}_k}{\vec{g}_{k-1} \cdot \vec{g}_{k-1}} \quad (3.11)$$

b_k étant le facteur pondérant les gradients courants et précédents, et g_k la direction de recherche.

Le déplacement se fera de façon à minimiser la fonction potentielle V suivant la relation :

$$V(\vec{X}_{k+1}) = V(\vec{X}_k + \lambda_k \vec{S}_k) \quad (3.12)$$

Le minimum sera déterminé en effectuant un test sur la norme du gradient :

$$\left| \left(\frac{\delta V(\vec{X})}{\delta \vec{X}} \right)_{X_{k+1}} \right| \leq \epsilon \quad (3.13)$$

où ϵ correspond au critère de convergence défini par l'utilisateur (généralement égal à 10^{-4} kcal mol⁻¹ Å⁻¹).

3.4.2 Dynamique moléculaire

Une fois les conformations de départ minimisées, elles pourront être utilisées en dynamique moléculaire. Cette méthode permettra de simuler et donc d'observer le comportement d'un système en fonction du temps et dans un environnement donné.

3.4.2.1 Dynamique Newtonienne

La dynamique moléculaire est basée sur l'hypothèse que les différentes configurations du système peuvent être générées par la résolution dans le temps des équations de mécanique classique de Newton. Le résultat de cette intégration donne les trajectoires (positions) et les vitesses des atomes au cours du temps.

L'algorithme va permettre de calculer l'accélération s'appliquant sur chaque atome i à partir de l'équation de la seconde loi de Newton

$$F_i = m_i A_i \quad (3.14)$$

où F_i est la force, A_i l'accélération et m_i la masse de l'atome i .

Les forces seront alors calculées en dérivant la fonction d'énergie potentielle par rapport aux coordonnées cartésiennes de chaque atome :

$$F_i = -\frac{\partial U}{\partial r_i} \quad (3.15)$$

Les équations du mouvement sont déterministes. Ainsi, si l'on connaît les coordonnées et les vitesses initiales, les vitesses et coordonnées à un autre pas peuvent donc être calculées.

La méthode standard de résolution numérique des équations du mouvement est une méthode par différences finies, dont l'idée générale s'exprime de la façon suivante : à partir des coordonnées et des vitesses initiales au temps t , les positions et les vitesses sont calculées, $t + \delta t$ où δt va représenter le pas d'intégration. Ce pas d'intégration va dépendre de la méthode d'intégration et du système lui-même. Le choix de la méthode d'intégration va donc être important.

3.4.2.2 Dynamique de Langevin

La dynamique de Langevin va permettre de coupler le mouvement des atomes à un bain thermostatique via l'inclusion de termes stochastiques dans l'équation de Newton utilisée pour la DM classique qui est déterministe.

$$m_i \dot{V}_i = -\frac{\partial U}{\partial r_i} - m_i \gamma_i V_i(t) + \mathcal{R}_i(t) \quad (3.16)$$

avec $\dot{V}_i = A_i = dV_i/dt$, γ_i représentant une fréquence de collision qui peut être assimilée à un terme de friction dissipatif s'opposant au mouvement de l'atome i puisqu'il sera relié au coefficient de friction ζ_i de par la relation $\gamma_i = \zeta_i/m_i$. Le terme $\mathcal{R}_i(t)$ correspond à une force aléatoire, décorrélée dans le temps, due aux collisions associées aux mouvements browniens des particules environnantes vérifiant la relation $\langle \mathcal{R}_i(t) \rangle = 0$.

Historiquement, la dynamique de Langevin a été utilisée comme une alternative à la DM en solvation explicite. Le système est séparé en deux sphères : une sphère interne dans laquelle les atomes se propagent par des méthodes de DM classiques en solvation explicite, et une seconde sphère externe, où la solvation est implicite, et où la dynamique de Langevin sera utilisée pour représenter les mouvements des atomes. Dans cette sphère, l'effet du solvant est représenté par les forces de friction correspondant au second terme de l'expression (3.16) ainsi que par la force stochastique $\mathcal{R}_i(t)$.

La dynamique de Langevin pourra cependant être utilisée en solvation explicite comme méthode de contrôle de la température du système.

3.4.2.3 Méthodes d'intégrations

Il existe différents algorithmes d'intégration des équations du mouvement au cours du temps dont notamment ceux de *Verlet*, *leap frog* ou encore *Predictor-Corrector*. Le pas d'intégration δt devra être choisi de façon à être suffisamment petit pour que les forces interatomiques puissent être considérées constantes. Dès lors, les méthodes analytiques de résolution des équations de Newton supposeront que les positions des atomes et des propriétés dynamiques du système pourront être développées en séries de Taylor :

$$\vec{X}(t + \delta t) = \vec{X}(t) + \delta t \vec{V}(t) + \frac{1}{2} \delta t^2 \vec{A}(t) + \frac{1}{6} \delta t^3 \vec{B}(t) \dots \quad (3.17)$$

$$\vec{V}(t + \delta t) = \vec{V}(t) + \delta t \vec{A}(t) + \frac{1}{2} \delta t^2 \vec{B}(t) + \frac{1}{6} \delta t^3 \vec{C}(t) \dots \quad (3.18)$$

$$\vec{A}(t + \delta t) = \vec{A}(t) + \delta t \vec{B}(t) + \frac{1}{2} \delta t^2 \vec{C}(t) + \dots \quad (3.19)$$

où X est le vecteur des positions, V est la vitesse (la première dérivée de la position en respect du temps), A correspond à l'accélération (la seconde dérivée), B est la troisième, etc.

Ensuite, les algorithmes diffèrent au niveau de l'approximation utilisée pour tronquer et évaluer le développement de $X(t + \delta t)$. La méthode la plus utilisée pour intégrer les équations du mouvement en DM est l'*algorithme de Verlet* [271]. Cet algorithme va utiliser les positions et les accélérations au temps t , ainsi que les positions au temps $(t - \delta t)$ pour calculer les positions au temps $(t + \delta t)$. Il en résultera donc les relations suivantes :

$$\vec{X}_i(t + \delta t) = \vec{X}_i(t) + \vec{V}_i \delta t + \frac{1}{2} \delta t^2 \vec{A}_i(t) + \dots \quad (3.20)$$

$$\vec{X}_i(t - \delta t) = \vec{X}_i(t) - \vec{V}_i \delta t + \frac{1}{2} \delta t^2 \vec{A}_i(t) - \dots \quad (3.21)$$

où V_i et A_i sont respectivement les composantes de la vitesse et de l'accélération.

L'addition de ces deux équations donnera :

$$\vec{X}_i(t + \delta t) = 2\vec{X}_i(t) - \vec{X}_i(t - \delta t) + \delta t^2 A(t) \quad (3.22)$$

Relation dans laquelle la vitesse est absente. Une des solutions va être de diviser la différence de positions aux temps $(t + \delta t)$ et $(t - \delta t)$ par $2\delta t$ et ainsi obtenir :

$$\vec{V}_i(t) = \left[\frac{\vec{X}_i(t + \delta t) - \vec{X}_i(t - \delta t)}{\delta t} \right] \quad (3.23)$$

L'initialisation de cet algorithme va nécessiter l'attribution à chaque atome de vitesses aléatoires à l'instant t_0 selon une distribution de Maxwell-Boltzmann à la température initiale souhaitée.

Le temps de calcul sera directement lié au pas d'intégration δt . En effet, à temps de simulation constant, plus celui-ci sera grand, plus le temps de calcul sera court. Cependant, le pas d'intégration doit rester largement inférieur (environ 1/10) à la période de vibration la plus faible du système (en général les vibrations des liaisons X-H qui sont d'environ 10 fs), soit une valeur de δt de l'ordre de 1 fs.

3.4.2.4 Algorithme SHAKE

Comme nous l'avons vu dans le paragraphe précédent, le pas d'intégration δt est limité par les vibrations internes de plus hautes fréquences. En limitant ces vibrations souvent sans importance (telles les vibrations de type X-H) pour les mouvements globaux, il sera possible d'augmenter le pas de simulation et ce, sans affecter le résultat de DM. Il existe de nombreux algorithmes (SHAKE, RATTLE,

SETTLE ...), permettant d'appliquer des contraintes de distance. Pour nos simulations, nous avons utilisé l'algorithme SHAKE [272] qui est l'algorithme le plus souvent utilisé en DM. Cette méthode permettra d'appliquer une contrainte de manière itérative sur une série de N_c paires d'atomes après chaque pas d'intégration, ce qui permettra de limiter les variations de longueur de liaison dans la limite d'un critère de convergence choisi.

L'utilisation de cette méthode au cours d'une DM permettra d'augmenter le pas de simulation à 2 fs.

3.4.3 Ensembles de simulations

Un ensemble est une collection de tous les systèmes possibles ayant des états microscopiques différents tout en appartenant à un état macroscopique ou thermodynamique unique.

Il existe différents ensembles thermodynamiques dans lesquels les simulations de DM pourront être effectuées.

L'intégration des équations de Newton conserve l'énergie totale du système. De plus, si le volume V de la boîte et le nombre N d'atomes sont maintenus constants, on générera un ensemble dit *micro-canonique* (NVE). Dans ce cas, la température et la pression du système seront des grandeurs dérivées. Lorsque l'on effectue une dynamique moléculaire, il sera parfois appréciable de pouvoir contrôler la température et/ou la pression du système de façon indépendante.

En effet, théoriquement l'ensemble NVE semble le mieux adapté à la DM puisque par définition l'énergie totale doit être conservée au cours d'une simulation. Cependant, l'accumulation d'approximations et d'erreurs au cours des pas d'intégration seront à l'origine de dérives de cette énergie totale au cours du temps.

Pour pallier à cela, ainsi que pour se rapprocher des conditions expérimentales, en plus de l'ensemble NVE, deux autres ensembles sont communément utilisés en DM. Il s'agira des ensembles *canonique* (NVT) et *isotherme-isobare* (NPT).

L'ensemble canonique NVT va permettre un contrôle de la température T du système à volume V constant. Cet ensemble sera généralement utilisé durant les étapes de thermalisation du système. L'ensemble permettant de limiter les dérives durant les pas d'intégration sera l'ensemble isotherme-isobare NPT. Cet ensemble permettra donc, en plus de contrôler la température T , de gérer la pression P du système. Nos simulations ont donc été réalisées dans ces conditions.

D'autres ensembles permettant de contrôler d'autres grandeurs thermodynamiques sont également accessibles aux simulations de DM tel l'ensemble *isobare-isoenthalpie* (NPH) qui aura un nombre fixe d'atome N , pression P fixe et enthalpie H fixe, ou encore l'ensemble *grand canonique* (μ VT) caractérisé par un potentiel chimique μ fixe à volume et température constants.

3.4.4 Conditions initiales de simulations

L'initiation d'une DM va nécessiter de connaître les coordonnées et les vitesses initiales de tous les atomes du système au temps t_0 . Les coordonnées atomiques initiales seront généralement issues de données expérimentales obtenues par cristallographie ou RMN. Les vitesses initiales quant à elles, seront attribuées de façon aléatoire à tous les atomes selon une distribution de Maxwell-Boltzmann centrée sur la température T , ce qui permettra le calcul des densités de probabilité $f(\mathbb{V}_i)$.

La distribution initiale des vitesses n'est pas équilibrée. Une période d'équilibration du système sera donc nécessaire afin de stabiliser la simulation.

3.4.5 Dynamiques moléculaires à température et pression constantes

3.4.5.1 Contrôle de la température

Lors de simulations, le contrôle de la température sera important notamment vis à vis des systèmes biologiques où cette dernière sera essentielle à certains processus structurels tels le repliement des protéines ou autres.

En DM, le système pourra être maintenu à température constante grâce à différentes méthodes. La méthode la plus simple est celle du recalibrage des vitesses. Cette méthode va consister à utiliser le théorème d'équipartition selon lequel chaque degré de liberté d'un système correspond à une énergie cinétique moyenne de $k_B T/2$, ce qui pour un système à N atomes donnera :

$$\langle E_C \rangle = \frac{3}{2} N k_B T \quad (3.24)$$

où T représentera la température désirée du système et k_B la constante de Boltzmann.

Au cours d'une simulation, l'énergie cinétique instantanée $E_C(t)$ varie et s'éloigne de $\langle E_C \rangle$. Les vitesses seront donc périodiquement corrigées d'un facteur λ de façon à ce que $E_C(t)$ tende vers $\langle E_C \rangle$:

$$\frac{3}{2} k_B T = \frac{1}{2} m_i [\lambda V_i(t)]^2 \quad (3.25)$$

La correction des vitesses se fera ensuite par l'intermédiaire des températures :

$$\lambda = \sqrt{\frac{T}{T(t)}} \quad (3.26)$$

où $T(t)$ correspondra à la température instantanée.

Ce type de thermostat est appelé thermostat isocinétique. Cette méthode aura tendance à entraîner des sauts d'énergie totale du système qui ne sera alors plus conservée.

Pour générer l'ensemble de distribution le plus correct, le système pourra être couplé à un réservoir. Le couplage sera soit stochastique, soit déterministe. Les couplages déterministes permettent généralement de conserver les quantités et ainsi de fournir des simulations fiables. L'approche stochastique quant à elle, est plus facile à implémenter et permet d'augmenter la stabilité des dynamiques via l'introduction d'un terme de friction.

Le logiciel NAMD [273] que nous avons choisi pour nos DM, utilisera l'équation stochastique de Langevin pour générer une distribution de Boltzmann pour des simulations dans l'ensemble canonique NVT. L'équation de Langevin modifiera l'équation de Newton du mouvement en ajoutant deux termes représentant le couplage au réservoir. Il s'agira d'un terme de fluctuation de la force (dernier terme de l'expression 3.27) ainsi qu'un terme de friction représentant la dissipation de la force (2^{ème} terme de l'équation 3.27). L'équation de Langevin sera donc la suivante :

$$m_i A_i = F_i(X_i) - \gamma V_i + \sqrt{\frac{2\gamma k_B T}{m_i}} \mathcal{R}(t) \quad (3.27)$$

où m_i est la masse, A_i correspond à l'accélération, F_i est la force, X_i est la position, γ est le coefficient de friction, k_B correspond à la constante de Boltzmann, T à la température et $\mathcal{R}(t)$ représente un processus Gaussien aléatoire.

L'intégrateur implémenté dans NAMD est celui proposé par Brünger, Brooks et Karplus [274] qui est une adaptation de l'algorithme de Verlet pour l'équation de Langevin.

Pour les équations stochastiques du mouvement, la position et la vitesse des atomes deviennent des variables aléatoires tandis que l'évolution au cours du temps de la fonction de distribution correspondante sera décrite par l'équation déterministe de Fokker-Planck [275].

3.4.5.2 Contrôle de la pression

De la même manière qu'un système simulé peut être couplé à un réservoir thermique, on pourra utiliser un réservoir de pression afin de réguler la pression du système au cours d'une simulation. La plupart des expériences de biologie sont faites à pression quasi-constante, de ce fait si l'on veut se rapprocher des conditions *in vivo* lors de simulations, le contrôle de la pression pourra donc s'avérer indispensable.

Le contrôle de la pression s'effectuera généralement via une correction appliquée au volume global du système, et donc à la position des atomes.

Le logiciel NAMD [273] met à notre disposition différentes méthodes afin d'effectuer des simulations à pression constante, telles la méthode de Berendsen ou une méthode basée sur la méthode de Nosé-Hoover modifiée, couplée à une dynamique de Langevin pour contrôler les fluctuations du barostat (méthode nécessitant d'être couplée à un contrôle de la température pour simuler l'ensemble NPT). Nous parlerons ici, uniquement de cette seconde méthode utilisée dans nos travaux pour simuler l'ensemble isotherme-isobare.

L'algorithme Nosé-Hoover piston de Langevin implémenté dans NAMD et une combinaison de la méthode à pression constante de Nosé-Hoover telle qu'elle est décrite par Martyna et ses collaborateurs [276], avec un piston de contrôle des fluctuations basé sur la dynamique de Langevin comme décrit par Feller et ses collaborateurs [277].

Ses équations vont permettre d'introduire le volume comme une variable dynamique. Le système est couplé au volume \mathcal{V} de la boîte de simulation afin de symboliser l'action d'un piston barostatique. L'effet du mouvement du piston pourra être décrit par la variation de \mathcal{V} au cours du temps selon l'équation de Langevin :

$$\ddot{\mathcal{V}} = \frac{P(t) - P_{ref}}{\mathcal{M}_p} - \Gamma \dot{\mathcal{V}} + R_p(t) \quad (3.28)$$

où $\ddot{\mathcal{V}}$ représente l'accélération du volume due au mouvement du piston, $P(t)$ et P_{ref} sont respectivement la pression instantanée à l'instant t et la pression de référence, \mathcal{M}_p est assimilable à la masse du piston, Γ correspond à la fréquence de collision, $\dot{\mathcal{V}}$ représente la vitesse du volume \mathcal{V} et $R_p(t)$ représente une force stochastique.

La température T du bain thermostatique va ensuite intervenir dans le contrôle de la pression : $R_p(t)$ sera extrait d'une distribution gaussienne de moyenne nulle et de variance $2\Gamma k_B T \delta(t) / \mathcal{M}_p$. La pression instantanée $P(t)$, elle, sera dérivée du théorème du viriel : $P(t)\mathcal{V} = \frac{3}{2} N k_B T$, ce qui donnera :

$$P(t) = \frac{1}{\mathcal{V}} \left(N k_B T - \frac{1}{2} \sum_{i=1}^N r_i(t) \cdot F_i(t) \right) \quad (3.29)$$

avec $P(t)\mathcal{V} = \sum_{i=1}^N r_i(t) \cdot F_i(t)$, et $F_i(t)$ correspondant à l'ensemble des forces exercées sur l'atome i à l'instant t .

3.4.6 Conditions aux limites

Lors d'une simulation en solvant explicite dans une boîte, la taille finie du système sera à l'origine d'effets de bords qui engendrent des biais sur l'environnement des atomes se trouvant à la périphérie de la boîte. Pour éviter cela, la solution sera de s'assurer que l'environnement des atomes devienne indépendant de leur position dans la boîte en simulant l'aspect infini d'une solution. On recourt ainsi aux conditions périodiques aux limites. Cette méthode va permettre de répliquer un ensemble fini et relativement restreint de particules d'une boîte centrale dans les trois directions de l'espace de telle sorte que les forces agissant sur les composants du système correspondent à celles d'un système beaucoup plus grand (Figure 3.2). Les mouvements des atomes dans la boîte centrale seront reproduits dans toutes les boîtes images. De plus, le nombre de particules dans chaque boîte demeure constant au cours de la simulation. En effet, si une particule quitte la boîte centrale, elle sera immédiatement remplacée par une particule image de la boîte adjacente du côté opposé (Figure 3.2). Cette procédure revient donc à simuler l'existence d'un réseau cristallin infini où le comportement des particules serait identique d'une maille à l'autre.

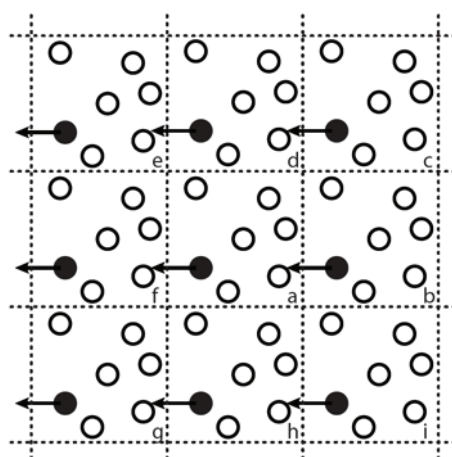


Figure 3.2 – Représentation en deux dimensions des conditions aux limites d'une boîte répliquée dans trois dimensions.

Dans ces conditions, les énergies sont calculées à partir des atomes situés dans la boîte centrale. Les interactions non liées des molécules de solvant proches des parois sont calculées à l'aide des molécules des boîtes adjacentes.

Cependant, il faudra s'assurer qu'un atome n'aura aucune interaction avec son image, et qu'une seule interaction avec un autre atome ou l'image de ce dernier.

3.4.7 Evaluation des interactions non liées

Simuler un système en DM va nécessiter le calcul de son énergie potentielle à chaque pas. Une grande partie du temps de calcul lors d'une dynamique va être dépensée à calculer les interactions non liées entre $N(N-1)/2$ paires d'atomes.

Au-delà d'une certaine distance entre deux atomes, le potentiel de Lennard-Jones, qui varie en R^{-6} , devient très faible, rendant ainsi les interactions de VdW négligeables.

Afin de réduire le nombre d'interactions non liées et ainsi gagner en temps de calcul, plusieurs méthodes ont été développées dont les plus utilisées sont l'utilisation de distances de troncatures non liées et de conventions de l'image minimum.

3.4.7.1 Troncature sphérique

Principe

Cette méthode va consister à définir un « *cutoff* » ou « *distance seuil* » R_C pour chaque atome i , au-delà de laquelle les interactions avec les autres atomes ne seront plus calculées.

Cette méthode va notamment s'appuyer sur la *convention de l'image minimale*, consistant à supposer que chaque atome i de la boîte centrale ne verra qu'une seule fois l'image de chaque autre atome présent dans le système. Les énergies

et/ou les forces seront donc calculées avec l'atome où l'image la plus proche de celui-ci.

Dans une boîte cubique, le rayon de troncature sera égal ou inférieur à la moitié de la longueur de la boîte. Cette approximation sera d'autant plus valable que la portée des interactions sera faible.

Cependant, l'utilisation d'une distance seuil n'aura pas une grande influence sur la réduction du temps de calcul puisqu'elle nécessitera l'établissement périodique d'une liste de toutes les distances interatomiques. En effet, calculer toutes les $N(N-1)/2$ distances coûte presque autant en temps de calcul que de calculer l'énergie elle-même.

Double cutoff

Afin de gagner du temps de calcul, la méthode du « *double cutoff* » [278] a été introduite. Cette méthode va consister à utiliser deux rayons de troncatures R_{C1} et R_{C2} . Tous les atomes j se trouvant à une distance inférieure à R_{C1} de l'atome i seront introduits dans une liste de paires d'atomes ou « *pair-list* » PL_i . Cette *pair-list* sera donc composée de l'ensemble des atomes j interagissant de manière non liée avec l'atome j . Les énergies d'interactions des atomes k correspondant à $R_{C1} < r_{ik} < R_{C2}$ seront stockées dans une variable E_k^{ILD} correspondant à l'énergie à longue distance (Figure 3.3).

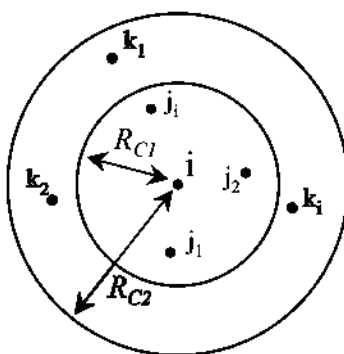


Figure 3.3 – Représentation schématique de la méthode du *double cutoff*

Lors de chaque pas t de DM, les énergies d'interactions non liées seront la somme d'une contribution à courte distance entre i et PL_i à l'instant t , et l'énergie E_k^{ILD} maintenue constante pendant plusieurs pas (PL_i et E_k^{ILD} étant renouvelés généralement tous les 10 à 50 pas).

Cette méthode de troncature sphérique, si elle reste acceptable dans l'estimation des interactions de VdW du fait de leur rapide décroissance lorsque les distances augmentent, s'avère inadaptée aux calculs des interactions électrostatiques qui elles diminueront lentement en fonction de l'inverse de la distance. En effet, l'introduction d'un *cutoff* va engendrer des brusques coupures dans les interactions et ainsi dans les calculs d'énergies et de forces. Ces effets pourront être atténués grâce notamment à l'application de fonctions de lissage utilisées comme facteurs multiplicatifs des termes non liés.

Fonctions de shift et de switch

La fonction de « *shift* » va agir sur l'ensemble de la surface d'énergie potentielle sous la forme :

$$\mathcal{S}_{shift}(r_{ij}, R_C) = \begin{cases} \left[1 - \left(\frac{r_{ij}}{R_C}\right)^2\right]^2 & \text{si } r_{ij} < R_C \\ 0 & \text{si } r_{ij} \geq R_C \end{cases} \quad (3.30)$$

Cette méthode modifiera l'ensemble du potentiel pour le faire tendre vers 0 à $r_{ij} = R_C$.

La fonction *switch*, quant à elle, va modifier le potentiel d'interaction au voisinage de R_C uniquement, et ce à partir d'une distance $R_{on} < R_C$:

$$\mathcal{S}_{switch}(r_{ij}, R_{on}, R_C) = \begin{cases} 1 & \text{si } r_{ij} < R_{on} \\ \frac{(R_C^2 - R_{ij}^2)^2 (R_C^2 + 2r_{ij}^2 - 3R_{on}^2)}{(R_C^2 - R_{on}^2)^3} & \text{si } R_{on} \leq r_{ij} < R_C \\ 0 & \text{si } r_{ij} \geq R_C \end{cases} \quad (3.31)$$

Ces deux méthodes vont éviter les discontinuités dans les potentiels, et vont donc autoriser des trajectoires stables sur de plus longues durées, mais ne résoudront pas le problème principal du calcul des interactions à longues distances.

3.4.7.2 Traitement des interactions à longues distances

Comme cité dans le paragraphe précédent, la discontinuité de l'énergie due à la distance du *cutoff* pourra être à l'origine d'artefacts dans les mouvements, les conformations ainsi que dans les énergies calculées. Ces artefacts pourront être corrigés via l'estimation des interactions au-delà du *cutoff* grâce notamment à la sommation d'Ewald [267,279].

Sommation d'Ewald

Cette méthode [280,281] conçue en 1921 dans le but d'évaluer les énergies d'interactions au sein de cristaux ioniques, va impliquer de traiter le système comme un quasi-cristal et d'effectuer les sommes par maille. Cette technique spécialement introduite pour calculer des interactions dans des conditions périodiques, suppose qu'une particule de la boîte (qui sera considérée comme une maille), interagit avec toutes les particules de la boîte ainsi qu'avec toutes leurs images se trouvant dans une matrice infinie de boîtes périodiques construite par

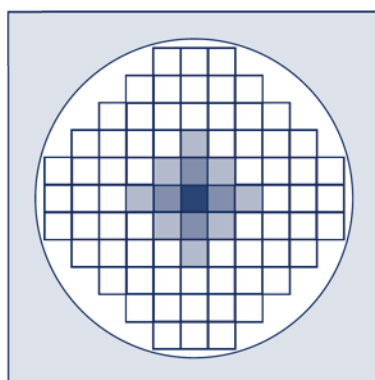


Figure 3.4 – Représentation d'un système en conditions périodiques construite avec la méthode d'Ewald. Adapté de Allen *et al.* [280].

translation de la maille initiale dans les trois directions x, y, z . A l'infini, le système global tendra en une forme quasi-sphérique (Figure 3.4).

Le potentiel électrostatique est décrit à l'intérieur de la boîte centrale de la manière suivante :

$$U_{el} = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (3.32)$$

qui en considérant les boîtes images donnera :

$$U_{el} = \frac{1}{2} \sum_n^* \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{4\pi\epsilon_0 |r_{ij} + n|} \quad (3.33)$$

où N est le nombre de particules dans chaque boîte, n est la distance avec la boîte centrale avec $n = n_x L_x + n_y L_y + n_z L_z$, L_x, L_y et L_z correspondant aux dimensions de la boîte, r_{ij} à la distance entre les particules i et j , et q_i, q_j à leurs charges. Le symbole $*$ indiquant que les termes correspondant à $i=j$ ne sont pas inclus lorsque $n = 0$.

Cependant cette équation converge très lentement, et seulement sous certaines conditions.

La sommation d'Ewald va donc consister à convertir cette équation lente à converger en deux termes qui convergeront rapidement en plus d'un ou deux termes constants.

Avec cette méthode, chaque charge sera considérée comme entourée d'une distribution de charge gaussienne neutralisante d'une magnitude égale mais de signe opposé (voir Figure 3.5).

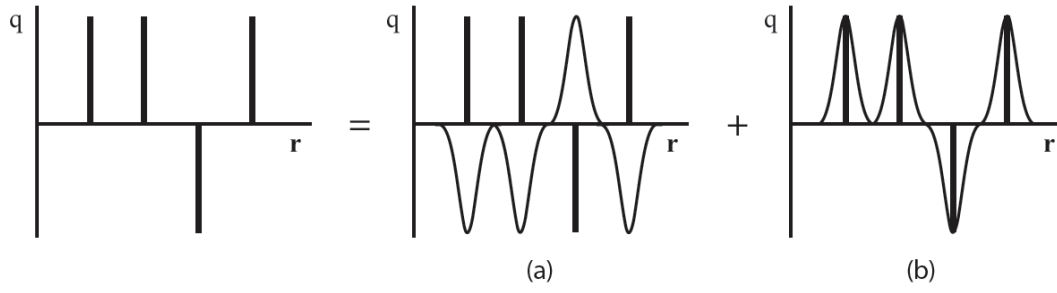


Figure 3.5 – Représentation schématique de la sommation d'Ewald dans un système unidimensionnel de charges ponctuelles. (a) Les charges ponctuelles (représentées par des traits verticaux) seront entourées par une distribution gaussienne de charges opposées calculées dans l'espace réel, auxquelles on ajoute (b) une distribution de charges calculées dans l'espace réciproque.

Le potentiel d'Ewald pourra donc s'exprimer sous une forme dont le préfacteur $1/4\pi\epsilon_0$ a été ignoré :

$$U_{Ewald} = U_{dir} + U_{rec} + U_{auto} + U_{surface} \quad (3.34)$$

avec U_{dir} représentant la somme directe qui est la somme sur l'espace réel, U_{rec} représente la somme sur l'espace réciproque, U_{auto} représente la somme des interactions de chaque gaussienne avec elle-même et un quatrième terme (dépendant du milieu) $U_{surface}$ représentant l'énergie de surface.

Ces termes pourront s'exprimer ainsi :

$$U_{dir} = \frac{1}{2} \sum_{i,j=1}^N \sum_n^* q_i q_j \frac{\text{erfc}(\beta |r_{ij} + n|)}{|r_{ij} + n|} \quad (3.35)$$

$$U_{rec} = \frac{1}{2\pi\mathcal{V}} \sum_{m \neq 0} \frac{\exp\left(\frac{-\pi^2 m^2}{\beta^2}\right)}{m^2} \left| \sum_{i=1}^N q_i \exp 2\pi i m \cdot r_i \right|^2 \quad (3.36)$$

$$U_{auto} = -\frac{\beta}{\sqrt{\pi}} \sum_{i=1}^N q_i^2 \quad (3.37)$$

$$U_{surface} = \frac{2\pi}{(2\epsilon_s + 1)\mathcal{V}} \left| \sum_{i=1}^N q_i r_i \right|^2 \quad (3.38)$$

avec q_i et r_i correspondant respectivement aux charges et coordonnées de l'atome i , β est un paramètre qui détermine le poids relatif des sommes directes et réciproques. \mathcal{V} représente le volume de la boîte de simulation, m est le vecteur réciproque et ϵ_s correspond à la constante diélectrique du milieu de la boîte de simulation.

La fonction d'erreur *erfc* (pour *complementary error function*) sera définie quant à elle par :

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt \quad (3.39)$$

Le terme U_{auto} sera une constante, tandis que le terme de l'énergie de surface $U_{surface}$ sera habituellement négligé en assumant que $\epsilon_s = \infty$ (ce qui correspondra à un milieu avec une permittivité relative infinie) ce qui est le cas pour l'eau qui est le solvant principalement utilisé en biologie pour recréer les milieux cellulaires.

Cependant, l'utilisation de cette méthode sera limitée par le calcul de la somme réciproque, dont le coût de calcul sera fonction de N^2 (voire $N^{3/2}$ pour les meilleurs algorithmes) avec N représentant le nombre de charges, ce qui empêchera son utilisation avec de gros systèmes.

Particle-mesh Ewald – PME

Particle-mesh Ewald (PME), est une méthode proposée par Tom Darden [237,238] afin d'améliorer le coût du calcul de la sommation réciproque. Cet algorithme va utiliser l'approche par troncature sphérique pour calculer la somme directe dans l'espace réel, tout en permettant de réduire le coût en temps de calcul nécessaire à l'évaluation du potentiel électrostatique dans l'espace réciproque en $N \log(N)$ grâce à l'emploi d'une transformée de Fourier rapide (*Fast Fourier Transforms*).

Pour cela, les charges ponctuelles du système sont interpolées sur une grille tridimensionnelle, la distribution des charges est alors calculée. L'étape suivante consistera à évaluer la contribution à longue distance du potentiel électrostatique dans l'espace réciproque, ce qui permettra d'interpoler en chaque point de la grille les forces et le potentiel liés à cette contribution. Finalement, grâce à leurs effets sur les charges ponctuelles, il sera possible de mettre à jour leurs coordonnées dans l'espace réel (voir Figure 3.6).

La méthode PME permettra donc un meilleur traitement des interactions électrostatiques à longues distances, tout en ayant un coût plus modeste en temps de calcul, rendant de ce fait, cette méthode particulièrement attractive dans le cadre des simulations de DM de macromolécules en solution, c'est pourquoi nous l'avons utilisée dans nos simulations.

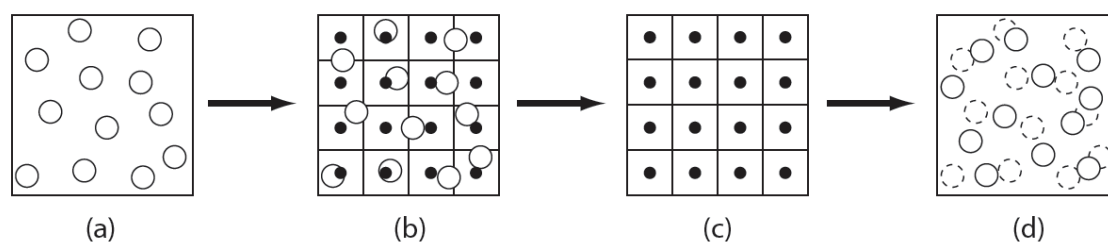


Figure 3.6 – Représentation des étapes de la méthode PME. (a) Système de charge dans un réseau bidimensionnel à l'instant t . (b) Interpolation des charges sur une grille bidimensionnelle. (c) Evaluation des forces et potentiels en chaque point de la grille. (d) Interpolation des charges avec mise à jour de leur position à l'instant $(t + \delta t)$.

3.5 Protocoles généraux de simulations

Comme déjà citées précédemment, les simulations de DM ont été réalisées avec le logiciel NAMD [273]. Ce logiciel gratuit présente l'avantage d'être hautement parallélisable, nous permettant de travailler avec plusieurs centaines de processeurs en même temps et ainsi gagner en temps de calcul. Au-delà de cet aspect pratique, tous les algorithmes et méthodes nécessaires aux simulations de DM (champ de forces AMBER, minimisation, Ewald, contrôle de température et pression, etc.) sont implémentés dans ce logiciel. De plus, NAMD nous permet d'implémenter facilement des contraintes généralisées nécessaire à nos simulations sous contraintes.

Les structures initiales utilisées pour nos simulations proviennent soit de structures cristallographiques issues de la PDB (pour notamment la structure de la daunomycine (DNM) intercalée entre les bases de l'ADN), soit de la base de Cambridge (pour notamment la structure de la DNM libre). Le logiciel de construction et de minimisation JUMNA [282] nous a également servi à construire des double hélices d'ADN.

La solvation des différents systèmes étudiés a été réalisée dans des boîtes cubiques de tailles variables (selon le système à étudier). Les systèmes ont ensuite été hydratés avec de l'eau TIP3P. Des ions (Na^+ et Cl^-) ont ensuite été ajoutés au système en quantités minimales pour neutraliser les systèmes ou en concentration physiologique.

Les deux étapes précédentes ont été réalisées avec le logiciel *LEaP* (issu de la suite de programme *AmberTools*). Ce logiciel nous a également permis de générer des fichiers de **topologies** contenant toutes les informations du système relatives aux champs de forces ou des charges.

Minimisations : Les systèmes moléculaires solvatés sont ensuite minimisés pendant un maximum de 5000 pas avec l'algorithme de minimisation fourni par NAMD en appliquant un critère de convergence de $10^{-4} \text{ kcal mol}^{-1} \cdot \text{\AA}^{-1}$.

Les simulations sont réalisées avec des conditions périodiques aux limites. L'évaluation des forces électrostatiques à longue distance est effectuée grâce à l'algorithme PME. Les simulations de DM sont effectuées en deux grandes étapes :

Simulation dans l'ensemble canonique NVT. Une fois les vitesses initiales assignées, le système est relaxé en plusieurs étapes dont les temps varieront en fonction des solutés. Au cours de ces étapes, la température du système sera maintenue à 300 K en utilisant la dynamique de Langevin. Les étapes de relaxation sont les suivantes : une étape où tous les atomes du soluté sont fixés en laissant libres toutes les molécules de solvant dans le but de laisser ces dernières se réarranger librement autour du soluté et ainsi éviter les instabilités du système. Les étapes suivantes vont consister à appliquer des contraintes harmoniques dont les forces vont en décroissant au long des étapes pour ainsi arriver à un système libre de toutes contraintes (sauf les distances X-H).

Simulation dans l'ensemble isotherme-isobare NPT. Le contrôle de la pression est appliqué lors de la dernière étape de simulation sous contraintes du soluté. Les simulations libres de toutes contraintes harmoniques sont donc ensuite poursuivies dans l'ensemble NPT pendant un certain nombre de nanosecondes afin d'assurer une stabilisation des solutés dans leur environnement. Ces DM sont effectuées en appliquant la méthode Nosé-Hoover couplée au piston de Langevin pour maintenir une pression constante à 1 atm et une température de 300K.

Nos systèmes ainsi relaxés sont donc prêts à servir de points de départ aux simulations nécessaires à nos travaux.

3.6 Energie libre

Depuis des dizaines d'années, l'estimation de l'énergie libre par des techniques de simulation moléculaire est au cœur de nombreuses études. L'énergie libre est souvent considérée comme étant la quantité la plus importante en thermodynamique. Elle permet d'expliquer l'existence et la plus ou moins grande facilité de mécanismes physicochimiques ainsi que la probabilité d'un système à adopter un état donné. La capacité à déterminer l'énergie libre à partir de simulations permet non seulement d'appréhender un processus au niveau atomique, mais également d'échantillonner des états d'un système difficilement observables de façon expérimentale. Finalement, si l'estimation de l'énergie libre à partir de simulations numériques est suffisamment précise, elle sera directement comparable avec les mesures expérimentales des propriétés thermodynamiques.

Dans un système en conditions NVT, l'énergie libre sera exprimée en termes d'énergie libre de Helmholtz, tandis que dans un système NPT, elle sera exprimée en termes d'énergie de Gibbs. Dans ce cas, l'énergie libre G d'un système sera donnée par la relation :

$$G = -\frac{1}{\beta} \ln Q \quad (3.40)$$

où $\beta = 1/k_B T$ (T correspondant à la température du système, et k_B à la constante de Boltzmann) et Q correspondant à la fonction de partition du système qui pourra être calculée via une intégrale sur tout l'espace des phases qui peut se présenter sous la forme :

$$Q = \int \exp[-\beta E(r)] d^N r \quad (3.41)$$

avec E correspondant à l'énergie potentielle et N au nombre de degrés de liberté du système.

Cependant, l'énergie libre (de même que les grandeurs associées telles l'entropie et les potentiels chimiques) est une quantité difficile à obtenir dans le cas de systèmes possédant de nombreux minima d'énergie locaux séparés par des faibles barrières d'énergie (tels les systèmes liquides ou les macromolécules flexibles). Les simulations de DM ou de Monte Carlo n'étant effectuées que sur des durées finies et obéissant à une statistique de Boltzmann, les parties de l'espace conformationnel situées à proximité d'un minimum seront bien échantillonnées, tandis que les régions de plus hautes énergies le seront bien moins. Pour des événements rares avec notamment des barrières d'énergie bien supérieures à $k_B T$, un échantillonnage direct sera hors d'atteinte des moyens de calcul actuels par plusieurs ordres de grandeur [283]. Ces régions de hautes énergies participent néanmoins à la fonction de partition du système et seront donc nécessaires à l'obtention d'un profil d'énergie libre complet.

Afin d'outrepasser ces limitations, différentes techniques ont été développées se basant sur des méthodes d'échantillonnage du système équilibré (telles l'intégration thermodynamique [284-286], la méthode de croissance lente [287,288] ou encore *l'umbrella sampling* [289,290]), non-équilibré (telle la méthode dite de croissance rapide [291-293]) ou encore des méthodes consistant à modifier la surface énergétique sur laquelle l'énergie libre sera calculée de façon à forcer une exploration plus étendue (telle la métadynamique [294]).

3.6.1 Calculs des différences d'énergie libre

L'obtention de l'énergie libre relative d'un système lors de simulations de DM (comme par exemple la différence de liaison de deux composés sur le même récepteur), revient à estimer la différence d'énergie libre (ΔG) entre deux états X et Y .

$$\Delta G = G_Y - G_X = -k_B T \ln \frac{Q_Y}{Q_X} \quad (3.42)$$

Afin d'estimer cette différence d'énergie libre, différentes approches pourront être utilisées. Nous décrirons brièvement ces différentes approches avant d'aborder les simulations effectuées lors de cette étude.

3.6.1.1 Perturbations thermodynamiques

Cette méthode, généralement attribuée à Zwanzig [295], est basée sur le fait que si l'on considère deux états X et Y , contenant N particules, interagissant respectivement selon les Hamiltoniens \mathcal{H}_X et \mathcal{H}_Y , la différence d'énergie libre entre les deux états pourra être définie par :

$$\Delta G = -k_B T \ln \left\{ \frac{\iint dp^N dr^N \exp[-\mathcal{H}_Y(p^N, r^N)/k_B T]}{\iint dp^N dr^N \exp[-\mathcal{H}_X(p^N, r^N)/k_B T]} \right\} \quad (3.43)$$

qui pourra se réécrire sous la forme :

$$\Delta G = -k_B T \ln \langle \exp -[\mathcal{H}_Y(p^N, r^N) - \mathcal{H}_X(p^N, r^N)]/k_B T \rangle_0 \quad (3.44)$$

L'indice 0 indique que l'ensemble des configurations représentatives de l'état X sont moyennées.

Cette approche du calcul de l'énergie libre consistera à définir les deux Hamiltoniens \mathcal{H}_X et \mathcal{H}_Y , puis à procéder à une simulation soit de l'état X , soit de l'état Y , et ainsi obtenir les ensembles moyens respectifs $\exp[-(\mathcal{H}_Y - \mathcal{H}_X)/k_B T]$ et $\exp[-(\mathcal{H}_X - \mathcal{H}_Y)/k_B T]$.

Cependant, dans les cas de recouvrement insuffisant des états X et Y dans l'espace des phases, l'échantillonnage de l'espace des phases au voisinage de Y lors de la simulation de X sera insuffisant (et inversement), rendant les valeurs de différence d'énergie libre calculées imprécises.

L'alternative la plus simple sera d'ajouter des états intermédiaires entre X et Y . Les termes des intermédiaires s'annulant, il sera donc possible d'en introduire au-

tant que nécessaire afin d'avoir un bon recouvrement et ainsi obtenir une bonne estimation de la différence d'énergie libre.

3.6.1.2 Intégrations thermodynamiques

Une autre méthode de calcul des différences d'énergie libre lors de simulations de DM, est l'intégration thermodynamique [284-286]. Cette approche va permettre de décrire la différence d'énergie libre par l'intégrale suivante :

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \mathcal{H}(p^N, r^N)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (3.45)$$

La résolution de l'équation précédente s'effectuera via une série de simulations correspondantes à des valeurs discrètes de λ entre 0 et 1 pour lesquelles on calculera la moyenne :

$$\left\langle \frac{\partial \mathcal{H}(p^N, r^N)}{\partial \lambda} \right\rangle_{\lambda} \quad (3.46)$$

Cette méthode permettra de considérer l'énergie libre en tant que somme de contributions des différents termes composant l'Hamiltonien.

3.6.1.3 Méthode de « croissance lente » - (« slow growth » method)

L'approche dite de « *croissance lente* » [287,288] consiste à calculer la différence d'énergie libre en faisant varier l'Hamiltonien en fonction d'une petite variable constante à chaque pas de simulation. Les Hamiltoniens aux pas de temps $t = i$ et $t = (i+1)$ seront donc presque égaux. Dans ce cas, la différence d'énergie libre sera représentée sous la forme :

$$\Delta G = \sum_{i=1}^N \sum_{\lambda=0}^1 (\mathcal{H}_{i+1} - \mathcal{H}_i) \quad (3.47)$$

Cette méthode sera cependant limitée, d'une part, par la variation d'énergie potentielle à chaque pas, ne laissant ainsi jamais assez de temps au système pour s'adapter aux nouvelles valeurs de paramètres de couplages, et d'autre part, par l'impossibilité d'améliorer l'échantillonnage de l'espace des phases à une valeur donnée sans perdre les informations des parties suivantes du calcul. De ce fait, cette approche n'est plus très utilisée de nos jours.

Toutefois, l'utilisation de ces méthodes pour le calcul des différences d'énergie libre présentera certains points faibles. En effet, avec toutes ces approches, la qualité des résultats sera non seulement dépendante de la qualité de l'Hamiltonien (qui est fonction du champ de forces choisi pour la simulation), mais également fortement corrélée à la qualité d'échantillonnages de l'espace des phases. Ce dernier point étant renforcé par l'absence de test garantissant le bon recouvrement des phases au cours des simulations.

3.6.2 Potentiels de force moyenne - PMF

Les méthodes décrites précédemment permettront d'obtenir les changements d'énergie libre dans le cas de « mutations » chimiques. Il est cependant intéressant d'étudier les variations d'énergie libre selon certaines coordonnées inter- ou intramoléculaires, telles des distances interatomiques, des angles de torsions, etc.

La surface d'énergie libre $G(\xi)$ le long d'une coordonnée réactionnelle (ξ) choisie porte le nom de « *potentiel de force moyenne* » (PMF) plus connue sous le nom de « *potential of mean force* ». Au contraire des mutations qui se feront le plus souvent le long de coordonnées non physiques, le PMF est généralement calculé pour des procédés physiquement accessibles. Le plus souvent il sera défini sur des critères géométriques.

Après avoir défini la coordonnée réactionnelle, la distribution de probabilité du système le long de ξ pourra être calculée par intégration de tous les degrés de liberté excepté ξ (d'où le terme « force moyenne »), et prendra la forme :

$$Q(\xi) = \frac{\int \delta[\xi(r) - \xi] \exp[-\beta E] d^N r}{\int \exp[-\beta E] d^N r} \quad (3.48)$$

qui sera reliée au PMF le long de la coordonnée réactionnelle ξ par :

$$G(\xi) = -\frac{1}{\beta} \ln Q(\xi) \quad (3.49)$$

Les simulations de DM ne permettront pas de calculer les intégrales directes des espaces de phases utilisées dans les équations (3.41) et (3.48). Cependant, si nous sommes en présence d'un système où tous les points des espaces de phases sont visités durant une simulation, $Q(\xi)$ sera égale à la densité de probabilité $P(\xi)$ le long de ξ . De ce fait, l'obtention de cette densité de probabilité lors de DM pourra directement être reliée à $G(\xi)$.

Lors du calcul du PMF, le point de plus haute énergie du profil correspondra par conséquent à l'état de transition du processus, à partir duquel il sera donc possible de dériver des quantités cinétiques.

Diverses méthodes de calculs du PMF ont été proposées, la plus couramment utilisée étant l'« *échantillonnage parapluie* ».

3.6.3 Umbrella sampling

3.6.3.1 Description

La méthode dite de l'« *échantillonnage parapluie* », ou « *umbrella sampling* », fut introduite à la fin des années 1970 grâce aux travaux de Torrie et Valleau [289,290] dans le but d'outrepasser les problèmes d'échantillonnage en modifiant

la fonction potentielle et ainsi permettre un échantillonnage suffisant de conformations défavorables.

Un biais sous la forme d'un terme d'énergie additionnel sera appliqué au système afin d'assurer un échantillonnage suffisant le long d'une coordonnée réactionnelle (ξ). Ceci pourra être effectué en une seule simulation, ou bien en plusieurs simulations dont les distributions se recouvrent ; dans ce dernier cas on parlera de « fenêtres » de simulation.

Le terme d'énergie additionnel sera représenté par un potentiel de biais w_i à la i ème fenêtre qui dépendra uniquement de ξ et qui prendra généralement la forme :

$$w_i(\xi) = \frac{1}{2} K (\xi - \xi_i)^2 \quad (3.50)$$

où ξ_i correspond à la valeur référence de ξ pour la fenêtre i .

L'énergie potentielle « biaisée » modifiée pourra être représentée par :

$$E^b(r) = E^{nb}(r) + w_i(\xi) \quad (3.51)$$

avec E^{nb} correspondant au potentiel « non biaisé ».

En accord avec la relation (3.48), la densité de probabilité du système non biaisé sera sous la forme :

$$P_i^{nb}(\xi) = \frac{\int \delta[\xi'(r) - \xi] \exp[-\beta E(r)] d^N r}{\int \exp[-\beta E(r)] d^N r} \quad (3.52)$$

Les simulations de DM permettront quant à elles d'obtenir les distributions biaisées P_i^b le long de ξ :

$$P_i^b(\xi) = \frac{\int \delta[\xi'(r) - \xi] \exp \left\{ -\beta \left[E(r) + w_i(\xi'(r)) \right] \right\} d^N r}{\int \exp \left\{ -\beta \left[E(r) + w_i(\xi'(r)) \right] \right\} d^N r} \quad (3.53)$$

A partir de là, les deux distributions pourront être reliées :

$$P_i^{nb}(\xi) = P_i^b(\xi) \exp [\beta w_i(\xi)] \langle \exp[-\beta w_i(\xi)] \rangle \quad (3.54)$$

A partir de cette dernière équation pourra être déduit $G_i(\xi)$ [283] :

$$G_i^{nb}(\xi) = -\frac{1}{\beta} \ln P_i^b(\xi) - w_i(\xi) + F_i \quad (3.55)$$

Avec $F_i = -(1/\beta) \ln \langle \exp [-\beta w_i(\xi)] \rangle$ indépendant de la coordonnée réactionnelle, $P_i^b(\xi)$ sera obtenue à partir de la simulation du système biaisé, et $w_i(\xi)$ sera obtenu de façon analytique.

Par la suite, l'obtention du PMF global $G(\xi)$, qui sera la combinaison des $G_i(\xi)$ de chaque fenêtre le long de ξ , se fera via l'utilisation d'algorithmes spécifiques, parmi lesquels on pourra notamment citer : la « *Conventional splicing Method* » [296-299], la « *Multiple Histogram Method* » [300,301], « *Weighted Distribution Function* » [302] (WDF) ou encore « *Weighted Histogram Analysis Method* » [303] (WHAM) et l'« *Umbrella Integration* » [304] (UI).

Lors de nos travaux, nous avons exclusivement utilisé l'algorithme WHAM décrit ci-dessous.

3.6.3.2 Weighted Histogram Analysis Method – WHAM

Cet algorithme développé par Kumar en 1992 [303] est une extension du Multiple Histogram Method proposé par Ferrenberg permettant d'estimer la valeur F_i et de combiner les résultats issus de simulations successives se recouvrant en se basant sur le calcul d'un facteur de pondération optimal pour chaque distribution à partir des points communs. Cette méthode est devenue la méthode standard pour combiner les résultats issus de différentes fenêtres de simulation [305], et à ce titre elle est implémentée dans de nombreux packages de logiciels de simulation [306].

Afin de minimiser l'erreur statistique sur P^{nb} le facteur de pondération p_i va donc être introduit :

$$P^{nb}(\xi) = \sum_i^{N_{sim}} p_i(\xi) P_i^{nb}(\xi) \quad (3.56)$$

avec N_{sim} correspondant au nombre de fenêtres de simulations.

L'algorithme sera sous la forme :

$$P_i^{nb}(\xi) = \frac{\sum_{i=1}^{N_{sim}} n_i(X)}{\sum_{i=1}^{N_{sim}} N_i \exp([F_i - w_i(\xi)]\beta)} \quad (3.57)$$

avec le coefficient F_i ,

$$F_i = -1/\beta \ln \left\{ \sum_{\xi} P_i^{nb}(\xi) \exp[-w_i(\xi)\beta] \right\} \quad (3.58)$$

où N_{sim} représente le nombre de simulations, $n_i(\xi)$ correspond au nombre de points dans la fenêtre i .

L'ensemble des sommes portant sur le nombre total de fenêtres ainsi que la valeur d'énergie libre seront obtenues par résolution itérative auto-cohérente des deux équations précédentes.

3.6.3.3 Analyses statistiques

Afin de caractériser l'erreur statistique des courbes d'énergie libre que nous avons obtenues par *umbrella sampling* couplée avec WHAM, nous nous sommes basés sur les travaux de Zhu et Hummer parus en 2012 [305]. Dans leur étude, ils ont montré que l'erreur statistique cumulée pour un PMF $G(\xi)$ le long de la coordonnée réactionnelle ξ échantillonnée par une série de i fenêtres avec des poten-

tiels harmoniques de type $K(\xi - r_i)^2/2$ (centré sur $r_i = r_0 + i\Delta r$), pouvait être exprimée en tant que racine carrée de la variance de l'énergie libre estimée :

$$var[G(\xi)] \approx (K\Delta r)^2 \sum_{i=1}^{(\xi-r_0)/\Delta r} var(\bar{\xi}_i) \quad (3.59)$$

où $var(\bar{\xi}_i)$ correspondra au carré de l'erreur lors de l'estimation de la position moyenne de ξ pour la fenêtre de simulation i . Ceci pourra être directement obtenu à partir de moyennes de ξ sur des blocs extraits de la fenêtre correspondante ; chaque fenêtre de simulation sera divisée en n blocs de taille m tels que les moyennes de ξ obtenues pour chacun des blocs soient indépendantes les unes des autres.

$$var(\bar{\xi}_i) = \frac{1}{n(n-1)} \sum_{b=0}^{n-1} \left[\frac{1}{m} \sum_{a=bm+1}^{(b+1)m} \xi_{ia} - \bar{\xi}_i \right]^2 \quad (3.60)$$

Dans nos simulations, des valeurs de n comprises entre 5 et 10 ont montré les meilleurs résultats, nous avons donc choisi une valeur intermédiaire avec $n=8$ pour nos calculs d'erreur statistiques.

3.6.3.4 Métadynamique

Au cours de notre étude publiée en 2012 [307] (Fournie en Annexe B) nous avons caractérisé l'intercalation d'une molécule d'intérêt biologique par *umbrella sampling*. Notre collaboration avec Arnab Mukherjee nous a permis de compléter cette caractérisation par la technique de *métadynamique*.

Cette technique introduite par A. Laio et M. Parrinello en 2002 [294], va permettre lorsqu'elle sera couplée à la DM, d'obtenir une surface d'énergie libre d'un

système ayant de nombreuses barrières d'énergie pouvant être problématiques pour des techniques standard [308].

Dans le cas d'une dynamique classique à N atomes, l'espace réel à explorer sera de dimension $3N-6$, ce qui, dans le cas d'un grand nombre d'atomes, rendra les simulations extrêmement longues. La métadynamique aura pour but de se restreindre à un sous-espace de l'espace réel des coordonnées de dimension d constitué d'un nombre limité de degrés de liberté, plus communément nommées variables collectives, correspondant à des variables d'intérêts (comme par exemple des distances, des angles dièdres, des angles, etc.).

Cette méthode permet d'accélérer la simulation en ajoutant un potentiel de biais externe histoire-dépendant à l'Hamiltonien du système. Ce potentiel pourra être décrit comme une somme de gaussiennes répulsives déposée le long de la trajectoire du système dans l'espace de la variable collective empêchant ainsi le système de revisiter des configurations déjà échantillonnées précédemment. Dans le cas où le système se trouve dans un minimum local d'énergie, le puits de potentiel sera ainsi rempli petit à petit, obligeant finalement le système à en sortir.

Considérons S un ensemble de d fonctions de coordonnées atomiques R d'un système :

$$S(R) = (S_1(R), \dots, S_d(R)) \quad (3.61)$$

A un temps t , le potentiel de biais pourra être représenté par :

$$V_G(S, t) = \int_0^t dt' \omega \exp \left\{ - \sum_{i=1}^d \frac{[S_i(R) - S_i(R(t'))]^2}{2\sigma_i^2} \right\} \quad (3.62)$$

avec σ_i correspondant à la largeur de la gaussienne pour la i ème variable collective et ω représentant un taux d'énergie vérifiant $\omega = W/\tau_G$ avec W représentant la hauteur de la gaussienne, et où τ_G correspond à la fréquence d'ajout de la gaussienne.

Par la suite, une estimation de l'énergie libre du système $F(S)$ sera obtenue à partir du potentiel de biais :

$$V_G(S, t \rightarrow \infty) = -F(S) + C \quad (3.63)$$

où C est une constante et $F(S)$ correspondra à :

$$F(S) = \frac{1}{\beta} \ln \left\{ \int dR \delta[S - S(R)] e^{-\beta U(R)} \right\} \quad (3.64)$$

avec $U(R)$ correspondant à la fonction d'énergie potentielle.

L'un des points primordial de cette méthode, est la pertinence des variables collectives utilisées. En effet, afin de conserver une application efficace de la métadynamique, ces variables devront notamment rester limitées en nombre, et présenter des états initiaux et finaux distincts tout en décrivant tous les états intermédiaires importants. (Voir Barducci *et al.* [309] pour une revue).

Finalement, une attention particulière devra être portée à la convergence des simulations. Il est essentiel de contrôler la durée des simulations afin que le potentiel de biais, après avoir complètement rempli les puits d'énergie libre dans la gamme pertinente des coordonnées utilisées, ne pousse le système à visiter des valeurs non pertinentes de ces coordonnées [309,310]. Afin de pallier ce problème, on pourra (i) limiter autant que possible la gamme autorisée des coordonnées (en imposant par exemple des « murs » répulsifs le long de ces coordonnées), (ii) s'assurer qu'en fin de métadynamique on observe des transitions rapides et nombreuses entre des conformations normalement séparées par de hautes barrières d'énergie libre, ce qui indique que le système diffuse librement le long des coordonnées contraintes dont les puits ont donc été correctement compensés par le potentiel de biais.

3.7 Outils d'analyses

Au cours des travaux présentés ici, nous avons utilisé divers outils pour réaliser l'ensemble de nos analyses et étapes préliminaires. Ces outils se présentent sous la forme de logiciels distribués de façon gratuite pour l'essentiel d'entre eux, ainsi que d'algorithmes classiques d'analyses structurales. Dans cette partie, nous allons présenter brièvement certains logiciels, algorithmes et protocoles que nous avons utilisé tout au long de cette étude.

3.7.1 Analyses structurales

3.7.1.1 Logiciels d'analyse

L'essentiel des analyses structurales (telles les mesures de distances, d'angles, d'angles dièdres, analyses des liaisons hydrogène, RMSD...) effectuées au cours de ces travaux a été réalisé via l'utilisation du logiciel de visualisation et d'analyse structurale VMD [311] (pour *Visual Molecular Dynamics*) ainsi que de certains outils pré-implémentés au sein de ce dernier. Ce logiciel nous a également permis de créer des représentations de l'ensemble des structures de nos systèmes présenté par la suite.

Dans le cadre de la construction de certains de nos systèmes, et plus précisément des phases de "docking" (ou amarrage moléculaire), nous avons fait appel au logiciel de visualisation Chimera [312].

3.7.1.2 RMSD

Le RMSD (pour *Root Mean Square Deviation* ou écart quadratique moyen) permet de mesurer la différence (en Å) entre deux structures S_0 et S_1 d'une molé-

cule à N atomes. Les deux structures ayant été préalablement superposées au mieux. L'algorithme sera de la forme :

$$RMSD(S_0, S_1) = \sqrt{\frac{1}{N} \sum_{i=1}^N |\mathbf{r}_{i,1} - \mathbf{r}_{i,0}|^2} \quad (3.65)$$

avec $\mathbf{r}_{i,0}$ et $\mathbf{r}_{i,1}$ correspondant aux vecteurs positions de l'atome i pour les structures S_0 et S_1 respectivement.

La structure utilisée comme point de référence sera le plus souvent, une structure issue de la RMN ou de la cristallographie aux rayons X, ou encore la première simulation générée au cours d'une simulation. On pourra ainsi calculer l'évolution du RMSD le long d'une trajectoire de DM en évaluant successivement à chaque pas de temps, le RMSD de chaque conformation par rapport à S_0 .

3.7.1.3 RMSD-2D

Un RMSD-2D consiste à calculer le RMSD de chaque conformation S_i d'un ensemble de structures S_n (à l'intérieur d'une fenêtre de simulation par exemple) par rapport à toutes les structures de ce même ensemble S_n . Le résultat sera représenté sous la forme d'une carte de densité dont les couleurs seront fonction de la valeur du RMSD. Ce type d'analyse permettra dans certains cas de mettre en évidence une homogénéité structurale au sein d'une trajectoire de DM, ou au contraire, révéler la présence de sous-ensembles structuraux bien distincts.

3.7.1.4 Génération de structures moyennes

Au cours de nos travaux, et plus précisément dans le cadre de nos simulations sous contraintes, nous avons généré des structures représentatives (S_{rep}) pour chaque fenêtre de simulation (S_i) le long d'un chemin réactionnel.

Dans ce sens, nous avons utilisé le module *g_covar* fourni avec le logiciel de DM *Gromacs* [313], qui nous a permis pour une fenêtre de simulation donnée, d'aligner l'ensemble des structures par rapport à un ensemble d'atomes donnés (par rapport aux atomes de phosphore de l'ADN dans notre cas), puis de générer à partir de là une structure moyenne S_m . Cette structure générée pouvant potentiellement présenter certaines incohérences structurelles, elle est ensuite comparée par RMSD à toutes les structures de la fenêtre de simulation. La structure ayant la plus petite RMSD par rapport à S_m est ensuite sélectionnée en tant que structure représentative de la fenêtre de simulation.

3.7.2 Curves+

3.7.2.1 Principes

Afin d'analyser les déformations de l'ADN, l'évolution de ses paramètres, ou encore l'influence de la présence de molécules intercalantes au cours de nos simulations, nous avons utilisé le logiciel d'analyse conformationnelle *Curves+* [314]. Ce logiciel dédié à l'analyse conformationnelle des acides nucléiques est développé depuis la fin des années 1980 par Richard Lavery et ses collaborateurs [315,316]. Il nous permet d'obtenir une description d'ensemble de la conformation d'un fragment d'ADN, en respect des conventions de Cambridge et Tsukuba [15,16]. Ce programme nous donne la possibilité d'accéder aux paramètres hélicoïdaux, aux paramètres du squelette ainsi qu'à une analyse des sillons de la double hélice.

Plus précisément, *Curves+* va nous permettre de caractériser notre fragment d'ADN selon des paramètres intra-paires de bases, inter-paires de bases, ceux du squelette phosphaté (que nous avons déjà décrit dans le Chapitre 1), ainsi que la géométrie des bases par rapport à l'axe hélicoïdal en plus d'une description des sillons de l'ADN en termes de largeur et de profondeur. Tous ces paramètres vont brièvement être décrits ici (voir l'article référence de Lavery *et al.* [314] pour une description plus approfondie des différents paramètres).

Paramètres intra-paire de bases : ces paramètres, au nombre de six, nous donneront des indications quant à la position des paires de bases vis à vis d'un ADN Watson-Crick canonique (Figure 3.7). Ils peuvent être divisés en trois paramètres de translation (« *shear* », « *stretch* » et « *stagger* »), et trois paramètres basés sur la rotation des bases (« *buckle* », « *propeller* » et « *opening* »). Tous ces paramètres prendront des valeurs égales à zéro dans le cas d'un ADN canonique. Toutes autres valeurs pourront être interprétées comme une déformation de la double hélice.

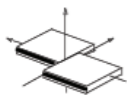

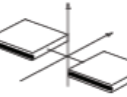

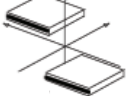
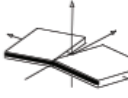
Translation		Rotation	
<i>Shear</i> (Cisaillement)		<i>Buckle</i> (Gauchissement)	
<i>Stretch</i> (Étirement)		<i>Propeller</i> (Torsion hélicoïdale)	
<i>Stagger</i> (Décalage)		<i>Opening</i> (Ouverture)	

Figure 3.7 – Représentation schématique des paramètres intra-paire de bases.

Paramètres inter-paires de bases : ces paramètres vont nous permettre de caractériser la position et donc les potentielles déformations de deux paires de bases successives. Ils seront également subdivisés en paramètres de translation (« *shift* », « *slide* » et « *rise* ») et paramètres de rotation (« *tilt* », « *roll* » et « *twist* ») (Voir Figure 3.8).

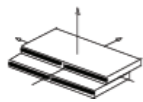
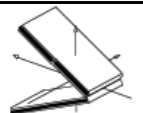
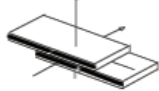
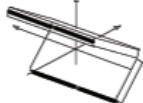
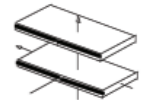
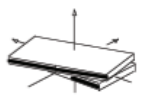
Translation		Rotation	
<i>Shift</i> (Glissement suivant x)		<i>Tilt</i> (Basculement)	
<i>Slide</i> (Glissement suivant y)		<i>Roll</i> (Roulis)	
<i>Rise</i> (Glissement suivant z)		<i>Twist</i> (Torsion)	

Figure 3.8 – Représentation schématique des différents paramètres inter-paires de bases.

Géométrie par rapport à l'axe hélicoïdal : ces paramètres nous permettent de caractériser les bases de l'ADN par rapport à l'axe hélicoïdal. Ils sont au nombre de quatre (Figure 3.9) : le « *Xdisp* » qui caractérise le mouvement des bases vers les sillons, le « *Ydisp* » qui permet de caractériser les mouvements perpendiculaires aux sillons, ainsi que l'« *inclination* » et le « *tip* ».

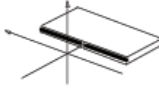
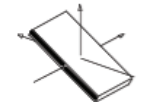
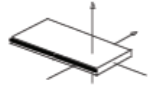

<i>Xdisp</i> (Déplacement selon x)		<i>Inclination</i> (Dévers)	
<i>Ydisp</i> (Déplacement selon y)		<i>Tip</i> (Inclinaison)	

Figure 3.9 – Représentation schématique des paramètres caractéristiques des bases de l'ADN par rapport à l'axe hélicoïdal.

Curves+ nous permet également de caractériser tous ces paramètres le long d'une trajectoire, ainsi que de produire une série de données de sortie graphique grâce au logiciel de complément *Canal* (pour *Curves+ analysis*).

3.7.2.2 Extension de Curves+ à la position du ligand

L'essentiel des systèmes que nous avons étudiés durant cette étude consiste à un complexe entre un dodécamère d'ADN complexé avec une molécule d'intérêt biologique (la daunomycine). De ce fait, nous avons ajouté une extension dans le programme afin de caractériser la position du ligand à tout moment par rapport à l'axe de l'ADN.

La localisation d'un ligand par rapport à la double hélice nécessite de définir un jeu d'axes basé sur une partie rigide du ligand. Dans le cas de la DNM, ce jeu implique uniquement des atomes appartenant aux cycles BCD (voir Figure 2.6) de la partie plane de la molécule (voir *supplementary material* de l'article fourni en Annexe B pour une description complète des différents référentiels dans le cas de la DNM).

Par la suite cette nouvelle option apportée au programme Curves+ nous permet de décrire la position du ligand grâce à trois paramètres de translation (« *L-shift* », « *L-slide* » et « *L-rise* »), et trois paramètres de rotations (« *L-tilt* », « *L-roll* » et « *L-twist* »), l'ensemble étant équivalent aux paramètres inter-paires de base standard de Curves+. En plus de cela, il sera également possible de calculer les translations (« *L-Tra* ») et rotations (« *L-Rot* ») totales du ligand par rapport à l'ADN.

CHAPITRE 4

Etude de la Flexibilité Interne de la Daunomycine

Une étude datant de 2008 menée par Mukherjee *et al.* [227] en collaboration avec notre laboratoire, sur laquelle nous reviendrons plus en détails dans les chapitres suivants, a notamment permis de proposer une description du chemin menant la DNM vers un site d'intercalation entre deux paires de bases de l'ADN. Dans ces travaux, le chemin d'intercalation de la DNM est essentiellement décrit en fonction des modifications qu'entraînera la DNM sur la structure de l'ADN sans considérer la flexibilité de la DNM en elle-même.

Comme nous l'avons vu au cours du Chapitre 2, le cycle A de la DNM présente une grande flexibilité et a fait l'objet de nombreuses études. Cette flexibilité de cycle pourra être caractérisée par l'étude de l'angle dièdre C7-C8-C9-C10 ou angle « τ ». Ce dièdre pourra adopter deux conformations préférentielles 9H_8 ($\sim 50^\circ$) et 8H_9 ($\sim -60^\circ$), plaçant respectivement le groupement acétyle en position équatoriale ou en position axiale (Figure 4.1). Au cours de notre étude, il sera indifféremment fait référence à la première conformation en tant que conformation « 9H_8 » ou « Eq », et à la seconde conformation en tant que structure « 8H_9 » ou « Ax ».



Figure 4.1 – Représentation des deux conformations principales du cycle A de la DNM. Conformation 9H_8 /Eq à gauche et conformation 8H_9 /Ax à droite.

Des analyses complémentaires des structures issues des travaux de Mukherjee *et al.*, ont pu mettre en évidence des transitions de conformation du cycle A le long du chemin de séparation des deux partenaires. Sachant que le processus d'intercalation nécessite une action collaborative des deux partenaires, il semble donc raisonnable de penser que ce phénomène de transition conformationnelle pourra également jouer un rôle important dans le processus de reconnaissance.

Partant de cette considération, l'étude de la flexibilité de la DNM dans différents environnements est une suite logique vers une meilleure compréhension du processus d'intercalation.

Dans ce sens, nos premiers travaux ont visé à caractériser la flexibilité interne du cycle A de la DNM, par dynamique moléculaire en la simulant libre et intercalée entre deux paires de bases. Afin d'être cohérents avec les travaux initiaux de Mukherjee, nous avons également simulé le complexe avec le ligand lié au petit sillon de l'ADN.

Dans un premier temps, nous nous sommes concentrés sur l'évolution de l'angle dièdre τ dans différentes conditions de simulation. Dans un deuxième temps, nous nous sommes intéressés à l'influence des transitions du cycle A en reproduisant nous-mêmes ces transitions de façon contrôlée par *umbrella sampling*.

Au cours de cette étude, notre attention s'est également portée sur la position spatiale du sucre daunosamine, et plus particulièrement sur son groupement ammonium. En effet, cet ammonium porteur de la charge positive du ligand, est reconnu pour jouer un rôle important dans les interactions entre les deux partenaires [217,317] ainsi que pour avoir une influence éventuelle quant à la spécificité de reconnaissance de séquence [192].

Les travaux décrits dans [227], à l'instar de nombreuses simulations de DM ont été effectués dans des conditions « minimales », c'est-à-dire avec un nombre d'ions juste suffisants pour neutraliser les charges des molécules du soluté. Cependant, ces quantités d'ions ne reproduisent pas les conditions physiologiques, conditions qui par ailleurs sont indispensables au bon fonctionnement des cellules. Les molécules de notre système étant chargées, cette différence de concentration ionique pourra potentiellement influencer les conformations des différents partenaires.

Dans ce sens, nos simulations ont été effectuées dans deux conditions différentes :

- en concentration minimale de sels

- en concentration physiologique de sels (soit 150 mM de sels monovalents)

Ces simulations nous permettront ainsi de mettre en lumière une éventuelle influence de la concentration ionique sur chaque partenaire ainsi que sur le système dans son ensemble.

4.1 Construction des systèmes

Tous nos systèmes initiaux ont été construits, minimisés, relaxés et simulés en suivant le même schéma. L'étape de construction du système solvaté a été réalisée suivant les étapes décrites dans le Chapitre 3.

Dans toutes nos simulations, nous avons choisi d'utiliser le champ de forces AMBER99 avec la modification *parmbsc0* pour décrire l'ADN. Les paramètres de champs de forces pour la DNM étant quant à eux issus des champs de forces AMBER99 et GAFF.

Pour les simulations de la DNM libre, la configuration initiale que nous avons utilisée provient de la structure cristallographique résolue par Neidle et Taylor en 1977 [179]. Les charges RESP de la molécule ont été calculées via l'utilisation du module *antechamber* de *AMBER7* ainsi que *Gaussian03* [318] en utilisant la méthode Hartree-Fock avec une base 6-31G*.

La structure a ensuite été placée dans une boîte cubique puis hydratée avec de l'eau TIP3P dans laquelle des ions ont été ajoutés à différentes concentrations (les différents paramètres étant résumés par le Tableau 4.1).

Pour les structures où la DNM se trouve complexée à l'ADN, nos constructions ont été réalisées suivant le protocole présenté dans [227].

La structure utilisée pour simuler la DNM en position intercalée entre les bases de l'ADN est issue d'un cristal, à une résolution de 1,2Å, publié par Wang *et al* [181] en 1987 et déposé à la *Protein Data Bank* (PDB) sous l'identifiant 1D11. Cette structure est composée d'une répétition de quatre paires de bases d(ACGT)₂. Deux DNM y sont intercalées entre les deux paires de bases CpG que contient la

double hélice. Une partie de cette structure a été utilisée comme point de départ à la construction d'un dodécamère dont nous nous sommes servis pour toutes nos simulations. A partir du triplet (ACG), dont la géométrie a été conservée, quatre paires de bases de séquence d(CGCG)₂ ont été ajoutées avant, et cinq paires de bases d(TGCGC)₂ après, par l'intermédiaire du programme de construction et de minimisation de structures *JUMNA*. Nous avons ainsi formé le dodecamère de séquence d(GCGCACGTGCGC)₂ avec la DNM intercalée entre les bases C6 et G7 présenté par la Figure 4.2.

Cette structure est ensuite placée dans une boîte cubique hydratée avec de l'eau de TIP3P, et mise en présence d'ions à deux concentrations différentes. Les détails de ces constructions sont résumés par le Tableau 4.1.

Basés sur les hypothèses issues des travaux de Mukherjee [227] proposant que le chemin d'intercalation de la DNM passerait par une étape de liaison au petit sillon de l'ADN, nous avons également construit puis simulé le complexe avec la DNM liée au petit sillon de l'ADN.

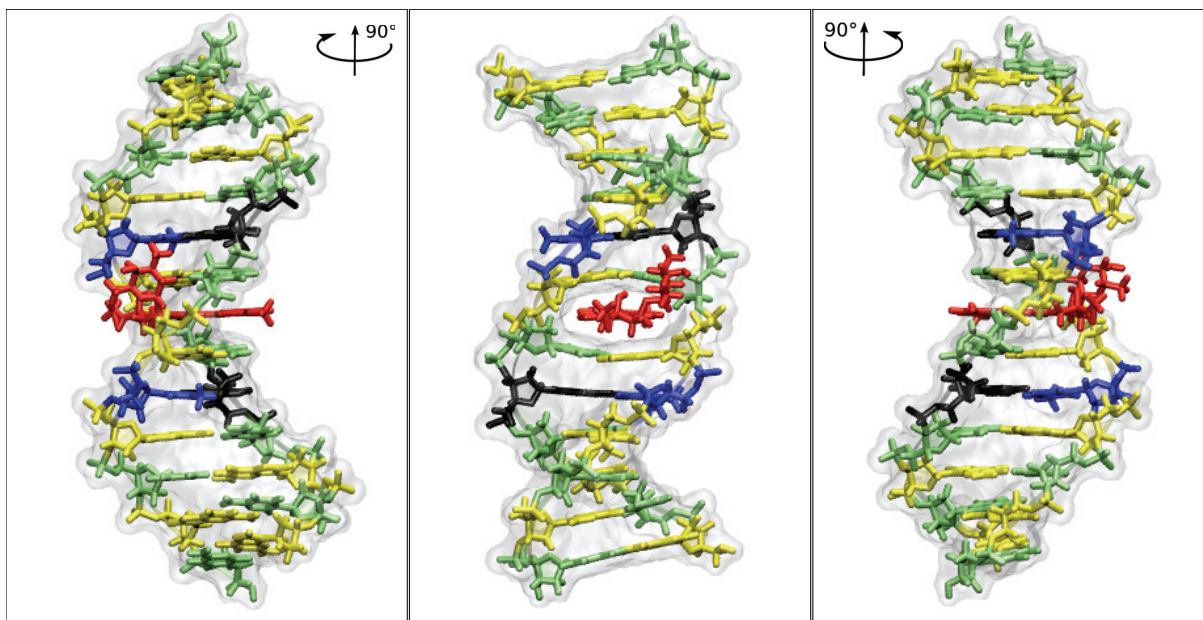


Figure 4.2 – Représentation de la structure de départ de la DNM intercalée entre deux paires de bases C6 et G7 de l'ADN.

Cette structure n'ayant jamais été observée expérimentalement, nous avons donc construit nous-même ce complexe. Pour cela, nous avons utilisé le programme *JUMNA* afin de construire une double hélice d'ADN-B de séquence identique à celle utilisée pour la simulation de la DNM intercalée. La DNM a ensuite été placée à la main dans le petit sillon à proximité du site d'intercalation C6pG7 (voir Figure 4.3). Le complexe ainsi obtenu étant ensuite minimisé avec le programme *JUMNA* pour réduire au maximum les contraintes stériques liées à l'assemblage.

La structure obtenue, est ensuite placée dans une boîte cubique puis solvatée avant d'être simulée dans les mêmes conditions que le complexe ADN-DNM intercalée (voir Tableau 4.1 pour un résumé des paramètres des boîtes de solvation).

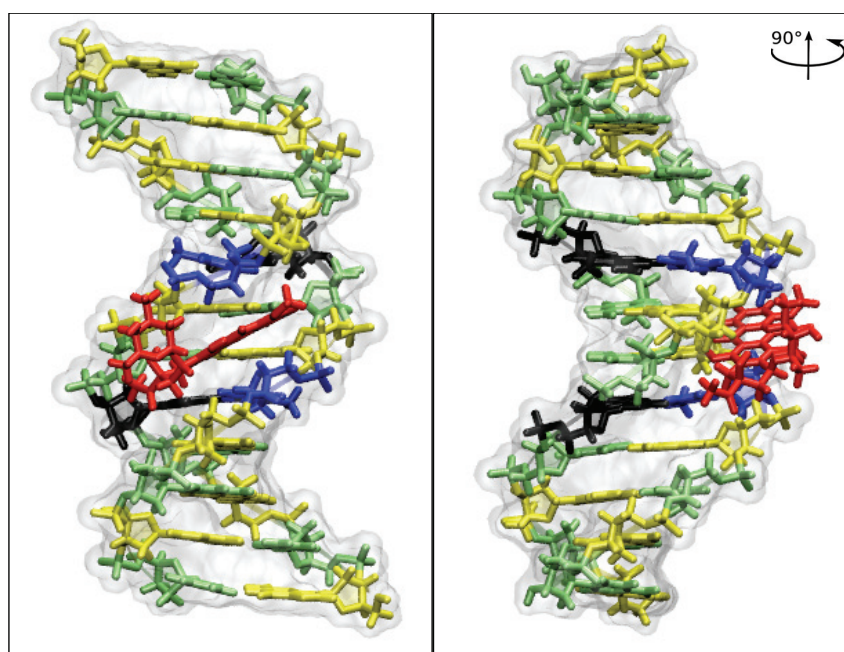


Figure 4.3 – Représentation de la structure de départ ayant servi aux simulations de la DNM liée au petit sillon de l'ADN.

Type de structure	Concentration en ions	Longueur du coté de la boîte (Å)	Volume de la boîte (Å ³)	Nombre de molécules d'eau	Nombre d'ions	
					Na+	Cl-
DNM libre	Minimale	37	50033	1097	0	1
	Physiologique (150mM)	41	69040	1603	6	1 + 6
DNM intercalée	Minimale	73	399256	10343	22	1
	Physiologique (150mM)	73	399256	10273	22 + 35	1 + 35
DNM liée au petit sillon	Minimale	69	336106	8639	22	1
	Physiologique (150mM)	69	336106	8587	22 + 29	2 + 29

Tableau 4.1 – Tableau récapitulatif des paramètres des boîtes de solvation des structures de départ.

4.2 Paramètres hélicoïdaux de l'ADN

Afin d'appréhender les effets de la DNM ou de la concentration en sels sur le double brin d'ADN, il est nécessaire d'obtenir une vue d'ensemble des paramètres hélicoïdaux de l'ADN libre. A partir d'une structure construite par l'intermédiaire du logiciel *JUMNA*, nous avons entrepris une caractérisation des paramètres d'un ADN-B canonique de même séquence que celle utilisée lors de nos travaux en présence de la DNM. Pour cela, l'ADN a été simulé pendant 50 ns dans des conditions identiques aux simulations des complexes décrits précédemment (même champ de forces, mêmes conditions d'hydratation) en présence de concentrations minimales et physiologiques de sels.

Dans nos constructions, la DNM étant soit intercalée entre les deux paires de bases centrales CpG, soit placée dans le petit sillon à proximité de ce même site d'intercalation, l'étude des déformations de l'ADN se limitera donc ici, au pas C6pG7 (les déformations des autres pas de la séquence d'ADN sont fournies en Annexe A).

La Figure 4.4 présente les paramètres hélicoïdaux de l'ADN pour le pas C6pG7 pour les simulations en conditions de sels minimales (représentées en bleu), ainsi que les simulations en conditions de sels physiologiques (représentées par les histogrammes rouges). Ces paramètres ont été déterminés à l'aide du programme Curves+ [314,316] respectant les conventions établies à Cambridge [15].

Dans le cas de notre séquence d'ADN, les deux concentrations ioniques que nous avons utilisées ne semblent avoir qu'une influence négligeable sur les paramètres hélicoïdaux et les largeurs des sillons au niveau du pas C6pG7 (voir Figure 4.4), mais également sur les autres pas de la double hélice présentés en Annexe A.

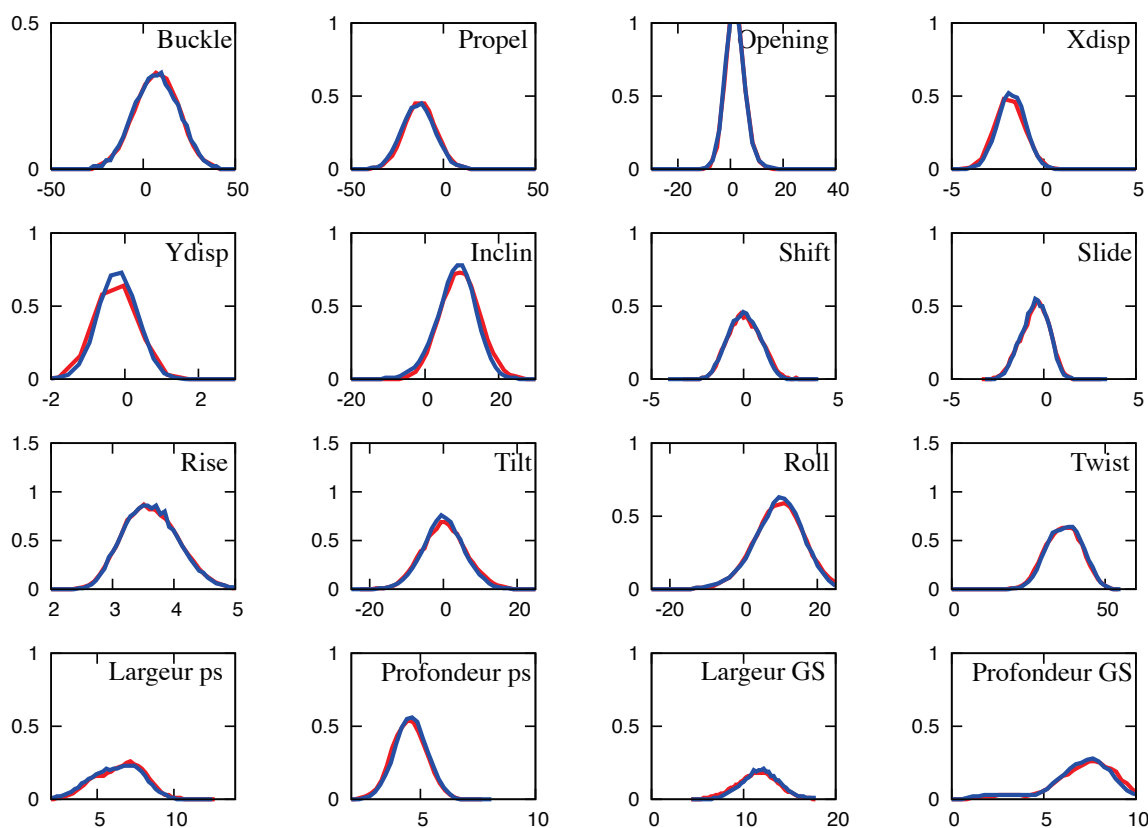


Figure 4.4 – Résumé des paramètres hélicoïdaux du pas C6pG7 d'un ADN-B canonique de séquence $d(GCGCACGTGCGC)_2$ après 50 ns de simulation en conditions minimales (bleu) et physiologiques (rouge) de concentrations en sels. *ps* correspondant à petit sillon, et *GS* à grand sillon.

4.3 Le cycle A de la daunomycine

4.3.1 Description du cycle A en fonction de son environnement

Comme décrit dans le Chapitre 2, différentes études ont décrit la flexibilité de la DNM et notamment du cycle A (voir paragraphe 2.3.2), dont le plissement pourra être déterminé par la mesure de l'angle de torsion τ .

D'après les structures cristallographiques issues de la littérature (résumées dans le Tableau 4.2), lorsque la DNM se trouve libre en milieu aqueux, le dièdre τ aura une valeur se trouvant entre 58° et 60° . Dans les complexes où la DNM est intercalée entre des paires de bases de l'ADN, l'angle τ prend des valeurs variant entre 40° et 66° avec plus de la moitié des structures cristallographiques publiées présentant des valeurs pour cet angle inférieures à 55° .

Les valeurs adoptées par le dièdre restent très proches des valeurs de la DNM libre, la présence de l'ADN a donc une influence limitée sur cet angle.

La position spatiale du sucre daunosamine de la DNM peut être étudiée via deux angles de torsions impliquant les atomes C7, O7 et C1' constituant la liaison entre le cycle A flexible et le sucre lui-même. Il s'agira des angles dièdre C7-O7-C1'-C2' (τ_1) et C6a-C7-O7-C1' (τ_2) (voir Figure 4.5). Ces deux angles affichent des valeurs respectives d'environ 165° et -115° dans le cas de la DNM libre (correspondant à des conformations *trans* et *gauche*-). Lorsque la DNM se trouve intercalée ils

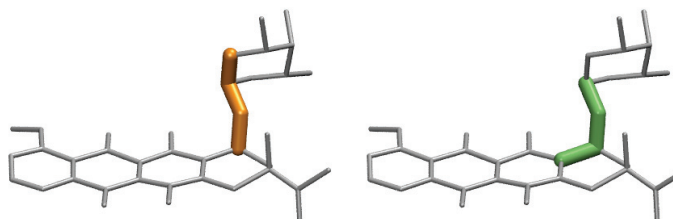


Figure 4.5 – Représentation des angles de torsion τ_1 (en orange) et τ_2 en vert permettant de caractériser la position spatiale du sucre par rapport au cycle conjugué de la DNM.

adoptent des valeurs situées entre 125° et 160° dont la majorité sont centrées autour de 140° pour τ_1 (soit une conformation proche de *trans*), et entre -150° et -120° avec la plupart des structures centrées autour de -145° pour le dièdre τ_2 (soit des conformations majoritairement *trans*).

Ces différences de valeurs en présence de l'ADN seront conséquentes de l'intercalation de la DNM. Dans ce cas, la partie conjuguée de la DNM se trouve profondément enfouie entre les bases de l'ADN, tandis que le sucre daunosamine reste dans le petit sillon, engendrant des contraintes stériques qui changeront la position du sucre par rapport au cycle rigide de la molécule.

Type de complexe	Source	ID PDB	Séquence (si existante)	Pas d'intercalation	Angle τ (°)	Angle τ_1 (°)	Angle τ_2 (°)	Ref
DNM intercalée (IC) dans l'ADN	PDB	1VTH	d(TGTACA)	CA	55	136,93	-124,85	[169]
	PDB	1VTI	d(TGTACA)	CA	66,38	126,53	-115,07	[169]
	PDB	110D	d(CGGCCG)	CG	39,43	161,73	-154,71	[184]
	PDB	152D	d(CGATCG)	CG	65,8	137,28	-140,3	[350]
	PDB	1D11	d(CGTACG)	CG	54,26	143,72	-144,22	[181]
	PDB	1DA0	d(CGATCG)	CG	52,94	148,71	-137,14	[186]
	PDB	1J02	d[(C(rG)ATCG)	CG	63,53	141,06	-144,81	[188]
DNM libre	Base de Cambridge	/	/	/	58	167,36	-113,9	[179]
	Base de Cambridge	/	/	/	60,21	163,27	-116,47	[178]
DNM IC ADN + Spermine	PDB	1D10	d(CGATCG)	CG	56	145,15	-152,25	[168]
DNM + Quadruplex d'ADN	PDB	100k	d(TGGGGT)	T	64,25	163,3	-137,37	[182]
					58,93	162,75	-137,61	
					56,66	164,63	-128,84	
DNM + TF	PDB	3F8F	/	/	-17,59	147,66	-156,25	[351]
DNM IC B-D-glucidé-ADN	PDB	308D	d(CGTACG)	CG	62,57	140,42	-137,29	[183]
					62,99	140,86	-138,66	
DNM IC/N-lié à l'ADN	PDB	4Z7D	d(CGC-(CH2-DNM)GCG	CG	58,4	143,69	-137,38	[187]
DNM IC/N-lié à l'ADN	PDB	1D33	d(CGC-(CH2-DNM)GCG)	CG	52,14	147,6	-138,64	[352]

Tableau 4.2 – Structures cristallographiques de la DNM se trouvant dans la littérature, classées selon le type de complexes qu'elles forment (soit intercalées dans l'ADN, seules, complexées avec d'autre type de macromolécules ainsi que liées de façon covalente en plus d'être intercalées (IC)), ainsi que leur identifiant PDB lorsqu'elles en ont un et la base de données d'où elles proviennent.

4.4 Etude des transitions conformationnelles du cycle A

Afin d'établir l'influence de l'environnement sur les conformations du cycle A, nous avons effectué une série de simulations visant à obtenir le profil d'énergie libre des transitions conformationnelles de l'angle dièdre τ par *umbrella sampling*.

Les simulations de dynamique moléculaire ont toutes été réalisées en conditions périodiques, les interactions électrostatiques à longues distances étant traitées via l'utilisation de l'algorithme Particle Mesh Ewald. L'algorithme SHAKE a été appliqué afin de contraindre la longueur des liaisons atomiques ce qui nous a permis d'utiliser des pas de simulations de 2 fs.

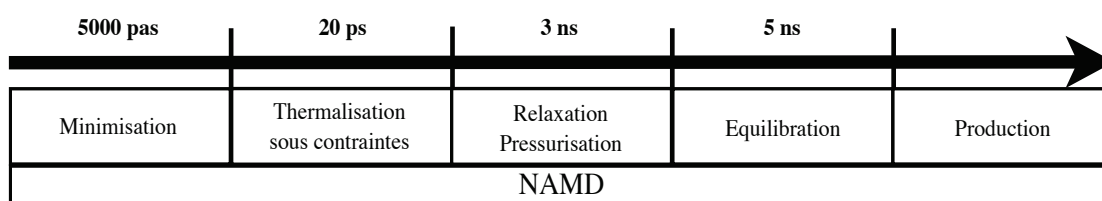


Schéma 4.1 – Description du processus précédant la simulation de dynamique moléculaire des complexes ADN/DNM.

Pour calculer l'énergie libre de la transition conformationnelle, nos simulations ont été effectuées en ajoutant une contrainte harmonique sur le dièdre τ . La coordonnée réactionnelle X , dans ce cas, est donc définie par la valeur de cet angle. Le long de cette coordonnée, nous avons appliqué un potentiel harmonique de la forme :

$$U = \frac{1}{2}k(X - X_0)^2 \quad (4.66)$$

où k est la constante de force et X_0 correspond au centre du potentiel harmonique.

A partir des structures décrites dans le paragraphe 4.2, préalablement équilibrées selon le Schéma 4.1, nous avons commencé une série de 31 simulations de 1 ns avec une constante de forces de l'ordre de 0,15 kcal mol⁻¹ deg⁻². La première de

ces simulations est effectuée en plaçant le centre du potentiel harmonique proche de la valeur de départ de l'angle dièdre pendant 1 ns. Les 30 simulations suivantes ont ensuite été réalisées en prenant la dernière structure de la simulation précédente comme point de départ, tout en bougeant le centre du potentiel harmonique de 5° par rapport à la simulation précédente. La première nanoseconde de simulation de chaque fenêtre de simulations a ensuite été augmentée jusqu'à l'obtention d'un total d'au moins 2 ns de simulation. Afin d'obtenir une vue d'ensemble de nos processus, les structures finales de chaque chemin de transition ont servi de point de départ à la construction de nouveaux chemins de transitions de sens opposé au premier.

4.4.1 DNM libre en milieu aqueux

Bien que la structure cristallographique utilisée comme point de départ à cette étude ait une conformation du cycle A de type Eq, après les étapes préliminaires de relaxation des systèmes, le cycle A se retrouve avec une conformation Ax, et ce, indépendamment de la concentration ionique. Les structures de départ utilisées pour simuler la transition du cycle A présentent donc des conformations axiales avec τ compris entre -58° et -55°. Par conséquent, les premiers centres de potentiel harmonique ont donc été fixés à -60°. Les chemins réactionnels ont été construits dans le sens de transitions Ax vers Eq (ou chemin AvE pour chemin de transitions de la conformation axiale vers la conformation équatoriale), puis dans un deuxième temps, dans le sens Eq vers Ax (que nous qualifierons de chemin EvA pour chemin de transitions de la conformation équatoriale vers la conformation axiale). Après la première nanoseconde de simulation servant à construire le chemin initial, l'échantillonnage de chaque fenêtre a été augmenté de 1 ns afin de permettre une convergence convenable des profils d'énergie libre.

La Figure 4.6 présente les profils d'énergie libre obtenus le long de la coordonnée réactionnelle. D'après ces profils, que ce soit en conditions minimales (Figure 4.6.a) ou physiologiques (Figure 4.6.b), les chemins de transitions de l'angle τ présentent deux minima locaux pour des valeurs de contraintes de -58° et

52°, correspondant respectivement aux conformations axiales et équatoriales du cycle A.

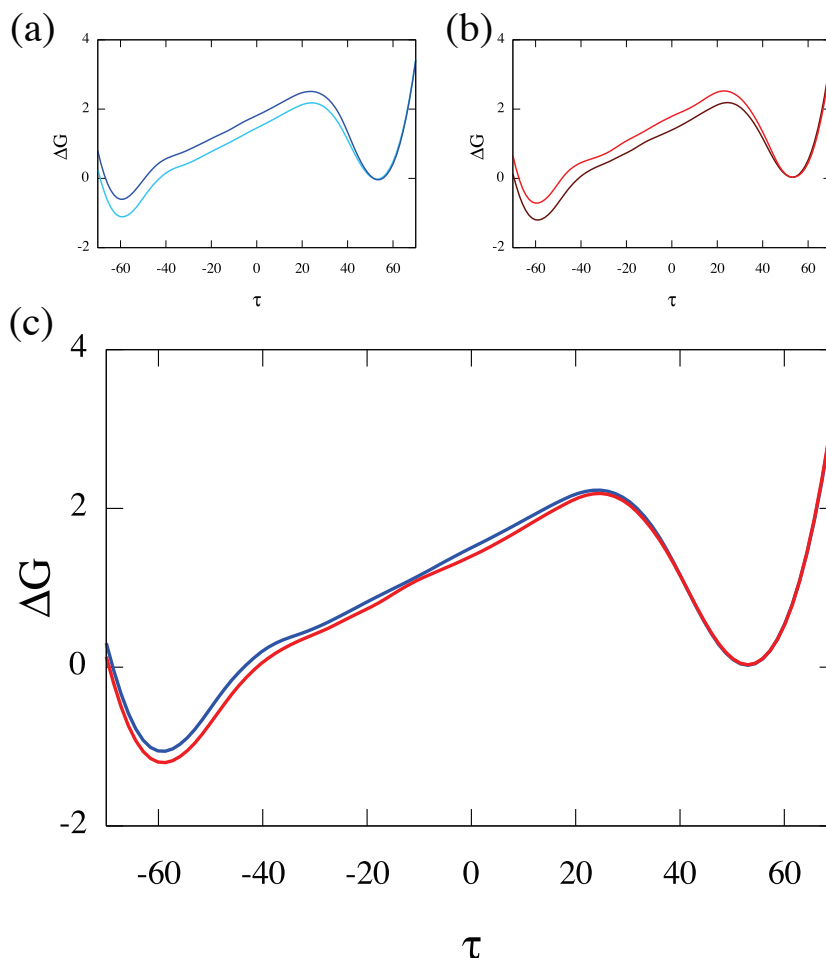


Figure 4.6 – Profils d'énergie libre le long du chemin de transition de l'angle τ . **a.** En conditions minimales de concentration en sels. Le chemin AvE est représenté en cyan, le chemin inverse est en bleu. **b.** En conditions physiologiques de sels (chemin AvE en marron et EvA en rouge). **c.** Représentation du chemin AvE en conditions minimales (bleu) et physiologiques (rouge) de concentration en sels.

Dans les deux conditions de simulation utilisées, les profils d'énergie libre présentent une très faible différence d'énergie ($< 0,5 \text{ kcal mol}^{-1}$) que le chemin réactionnel aille dans un sens ou dans l'autre. La conformation Ax présentant un minimum local plus bas en énergie (entre $0,5$ et $1,2 \text{ kcal mol}^{-1}$ selon le profil) que la conformation Eq.

Les profils d'énergie libre nous montrent donc que le cycle A de la DNM peut adopter deux conformations de stabilité similaire, séparées par une barrière d'énergie d'environ $3,5 \text{ kcal mol}^{-1}$. Ce résultat est cohérent avec les transitions conformationnelles du cycle A observées sur le chemin de sortie de la position intercalée.

Les structures cristallographiques de la DNM libre présentent généralement une conformation Eq du cycle A. L'étude de nos profils d'énergie libre montre néanmoins que la différence avec la structure Ax est faible, au moins pour les concentrations de sels que nous avons étudiées.

4.4.2 DNM intercalée

4.4.2.1 Etudes des chemins de transition du cycle A en conditions de sels minimales

Profils d'énergie libre

Comme attendu, les profils d'énergie libre (Figure 4.7) obtenus à partir des chemins de transitions EvA et AvE présentent tous deux deux minima locaux pour des valeurs du dièdre τ proche de 50° et -60° . Ces valeurs correspondent respectivement à des conformations de types Eq et Ax du cycle A.

Dans le sens de la transition EvA, contrairement aux simulations avec le ligand libre en milieu aqueux, le profil d'énergie libre (Figure 4.7 courbe bleue foncée) nous montre une différence d'énergie entre les minima de 6 kcal mol^{-1} en faveur de la conformation Eq. Les deux conformations étant séparées par un maximum d'énergie de 8 kcal mol^{-1} par rapport à la conformation de départ. La conformation Eq du cycle A est donc fortement stabilisée lorsque le ligand est intercalé. Ces résultats sont en accord avec les observations cristallographiques [171] lorsque la DNM est intercalée. Les différences de profil entre le ligand libre ou intercalé pourront s'expliquer par la présence de liaisons hydrogène à l'interface DNM-ADN sta-

bilisant les diverses conformations du cycle A lorsque la DNM est en présence de l'ADN. Nous reviendrons plus en détails sur ces liaisons hydrogène plus loin.

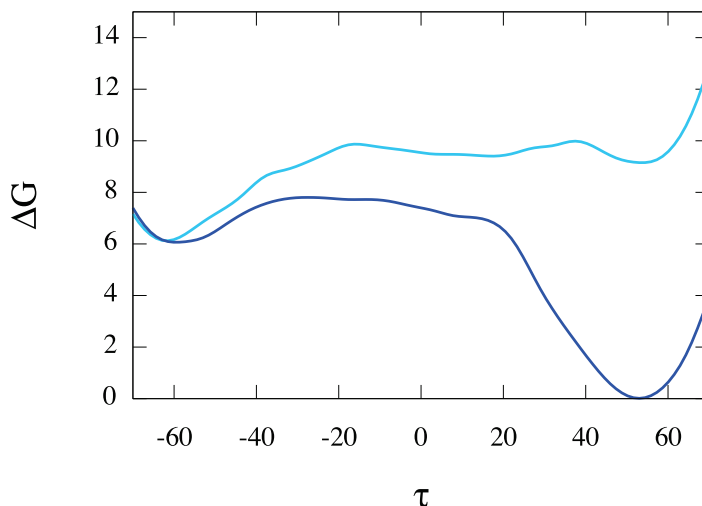


Figure 4.7 – Profils d'énergie libre (kcal mol⁻¹) des transitions du dièdre τ (en degrés) lorsque la DNM est intercalée entre deux paires de bases de l'ADN en conditions minimales, simulés dans le sens EvA (courbe bleue foncée) et AvE (courbe bleue claire).

Dans le cas du chemin AvE, le profil d'énergie libre obtenu (Figure 4.7 courbe cyan), s'avère de façon surprenante différent de celui obtenu pour le chemin EvA. Ici, la conformation axiale est plus stable que la conformation équatoriale avec une différence d'énergie de 3 kcal mol⁻¹ entre les deux conformations séparées par une barrière d'énergie de 4 kcal mol⁻¹.

Contrairement au cas où la DNM est simulée libre dans le milieu aqueux, les différences de profils en fonction du sens de simulation sont importantes. Ces différences sont en fait les conséquences de difficultés d'échantillonnage de tout l'espace conformationnel séparant les deux minima, qui se refléteront par des différences structurales. En effet, si l'on compare l'évolution des complexes le long des deux chemins, on peut observer de nombreuses différences tant au niveau des interactions à l'interface qu'au niveau des conformations de chacun des partenaires.

Interactions interfaciales

Le long de nos chemins de transition, la DNM crée préférentiellement des liaisons hydrogène stables via les oxygènes O11, O9 (correspondant à l'oxygène du groupement hydroxyle lié au cycle A) et O13 (correspondant à l'oxygène du groupement acétyle de la DNM). L'azote N3' du groupement ammonium porté par le sucre pourra quant à lui former des liaisons hydrogène avec les bases de l'ADN, mais également former des liaisons de type pont salin avec le squelette de la double hélice (Figure 4.8). Ici nous nous intéresserons uniquement aux interactions concernant les trois derniers atomes cités puisqu'ils se trouvent en relation directe avec le dièdre τ du cycle A.

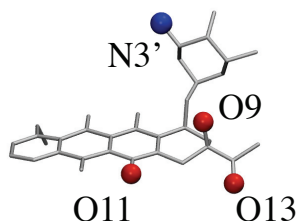


Figure 4.8 – Représentation des atomes de la DNM effectuant des liaisons hydrogène stables avec l'ADN le long des chemins de transition de la partie flexible du cycle A. Les oxygènes étant représentés en rouge et l'azote en bleu.

Le long du chemin EvA, pour des valeurs de $\tau > 20^\circ$, les structures présentent des interactions intermoléculaires sous forme de liaisons hydrogène via l'oxygène O9 (voir Figure 4.9). Cette liaison hydrogène aura pour effet de fortement favoriser la conformation Eq de la DNM, comme le montre le profil d'énergie libre sur lequel on peut observer un puits d'énergie corrélé avec la présence de telles liaisons. L'atome O9 étant directement lié au dièdre τ , ceci pourra donc expliquer sa forte influence sur la transition de l'angle de torsion.

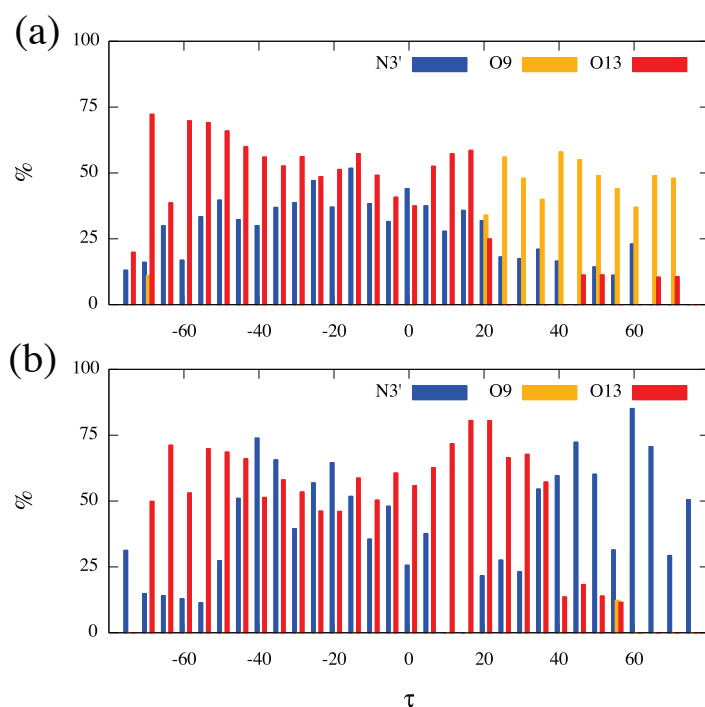


Figure 4.9 – Pourcentage de structure par fenêtre de simulation effectuant une liaison hydrogène impliquant l’azote N3’ (bleu) ou les oxygènes O9 (jaune) ou O13 (rouge). Dans le cas où un de ces atomes effectuerait des liaisons avec d’autres atomes lors de l’échantillonnage, seule la liaison étant la plus échantillonnée est représentée. **a.** Le long du chemin EvA. **b.** Le long du chemin de transition AvE.

La diminution de la valeur de τ de 20° à 15° s’accompagne de la disparition des interactions via l’atome O9 ainsi que d’une réorientation du sucre dans l’espace, illustrée par la transition du dièdre τ_1 (voir Figure 4.11 courbes bleues). Par la suite, toutes les structures issues des échantillonnages pour des valeurs de $\tau < 15^\circ$ présentent des interactions interfaciales plus ou moins stables, via des liaisons hydrogène impliquant l’oxygène O13, et l’azote N3’.

Le profil des liaisons hydrogène interfaciales le long du chemin AvE, est très proche de celui du chemin EvA pour des valeurs de $\tau < 10^\circ$. Ce résultat est en accord avec les profils d’énergie libre dans cette zone (même si ceux-ci présentent jusqu’à 2 kcal mol^{-1} de différence d’énergie, différence que l’on pourra imputer à la précision de la méthode ainsi qu’aux difficultés d’échantillonnages). Pour des valeurs de $\tau > 20^\circ$, le chemin AvE se différencie du chemin EvA par l’absence de

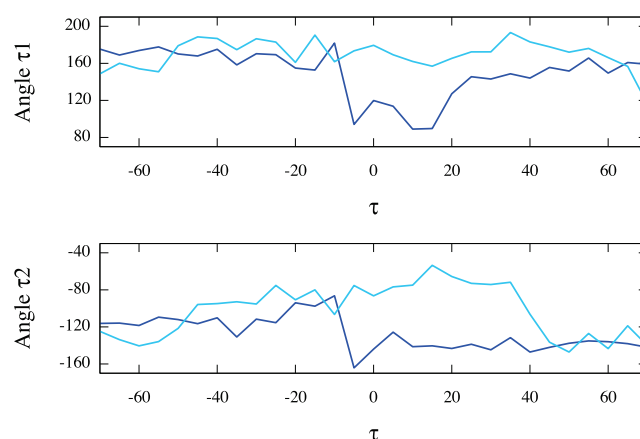


Figure 4.11 – Evolution des angles τ_1 (ou dièdre C7-O7-C1'-C2') et τ_2 (ou dièdre C6a-C7-O7-C1') exprimés en degrés, pour les structures représentatives de chaque fenêtre d'échantillonnage le long des chemins de transitions EvA (bleu) et AvE (cyan).

liaisons hydrogène impliquant l'oxygène 9. D'après la Figure 4.9 ces interactions sont cependant remplacées par des interactions via l'atome N3'. L'étude de ces interactions (basée sur la distance entre l'azote de la DNM et les oxygènes du squelette de l'ADN) illustrée par la Figure 4.10, nous montre qu'il s'agit en fait de ponts salins stables établis entre l'ammonium du ligand et le squelette phosphaté de l'ADN.

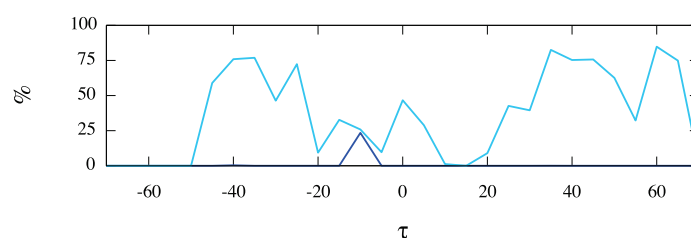


Figure 4.10 – Pourcentage de structures de chaque fenêtre de simulation effectuant un pont salin entre l'ammonium et le squelette de l'ADN le long des chemins EvA (en bleu) et AvE (en cyan).

La formation de ces ponts salins stables est le reflet d'une orientation différente du sucre daunosamine ainsi que de la DNM dans l'espace pour le chemin AvE par rapport au chemin EvA. Ce phénomène sera directement observable sur

les structures du complexe lorsque le cycle A est en position Eq (Figure 4.12), ainsi que par l'évolution des angles τ_1 et τ_2 , qui présentent des profils différents dans cette zone selon le chemin de transition considéré (voir Figure 4.11). Cette différence d'orientation spatiale du sucre, stabilisée par la formation d'une liaison type pont salin, aura pour conséquence de rapprocher l'oxygène O7 de l'hydroxyle lié en C9, favorisant de ce fait la formation d'une liaison hydrogène intramoléculaire (liaison déjà observée dans d'autres travaux expérimentaux [195,197,199]). L'oxygène O9 étant impliqué dans une liaison intramoléculaire, ses interactions avec l'ADN seront de ce fait très réduites.

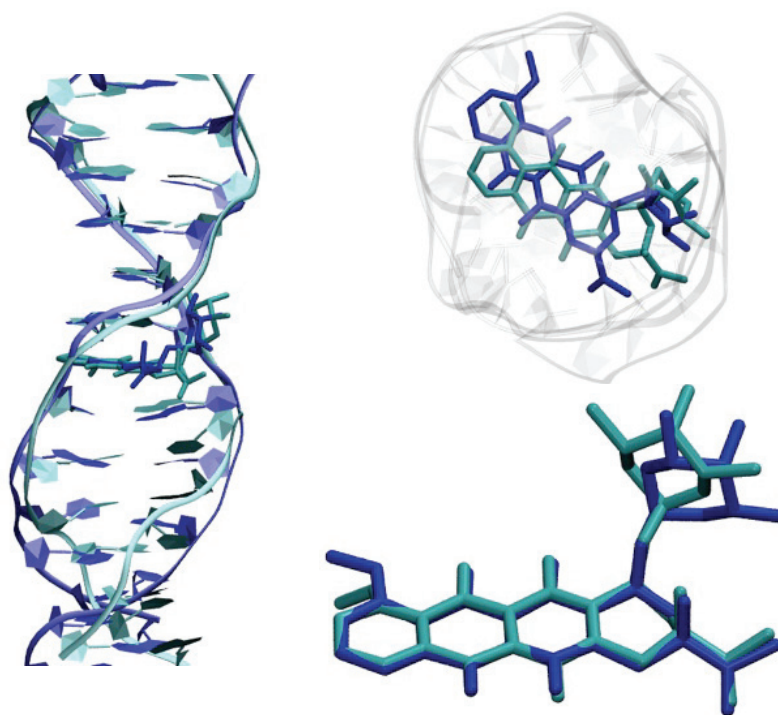


Figure 4.12 – Superposition des structures issues du chemin EvA (en bleu) et AvE (en cyan) lorsque l'angle τ se trouve à proximité de son minimum d'énergie en position Eq ($\sim 50^\circ$). Les structures sont superposées par rapport aux bases du site d'intercalation, vu du petit sillon (gauche) et vu du haut de la séquence (en haut à droite ; les molécules d'ADN sont représentées en transparence pour faciliter la visualisation). En bas à droite, représentation de la superposition des ligand. Les deux structures adoptent clairement deux positions différentes dans la poche d'intercalations ainsi que des orientations différentes de leur sucre dans l'espace.

4.4.2.2 Etude des chemins de transition en conditions de sels physiologiques

Profils d'énergie libre

Comme pour les simulations effectuées en conditions de sels minimales, les deux profils issus des chemins de transitions simulés en conditions physiologiques présentent une dépendance au sens de construction. En effet, le chemin EvA (Figure 4.13 courbe rouge) présente un profil d'énergie libre similaire au profil en conditions minimales mais avec une différence d'énergie entre les deux minima de 4 kcal mol⁻¹ en faveur de la conformation équatoriale. Le chemin AvE, quant à lui (Figure 4.13 courbe marron), présente deux minima de stabilités très proches, séparés par une barrière d'énergie de 4 kcal mol⁻¹.

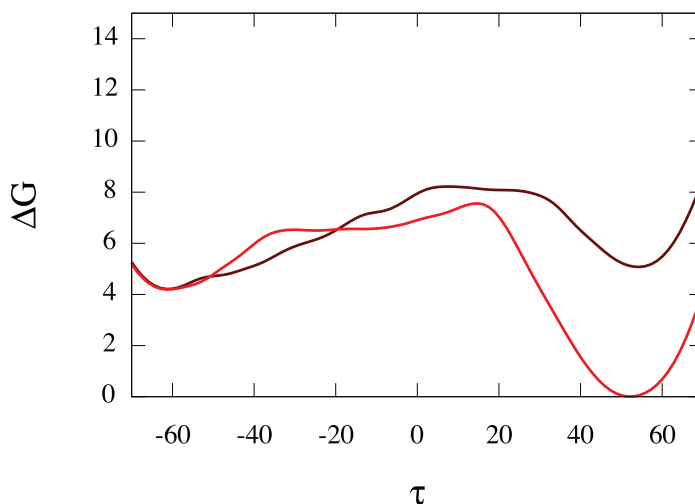


Figure 4.13 – Profils d'énergie libre (en kcal mol⁻¹) des transitions du dièdre τ (en degrés) lorsque la DNM est intercalée en conditions physiologiques, simulés dans le sens EvA (courbe rouge) et AvE (courbe marron).

Interactions interfaciales

Les structures du chemin EvA en conditions physiologiques présentent des interactions à l'interface entre les deux partenaires proches de celles observées pour les simulations en conditions minimales. Notamment, la présence de liaisons hydrogène stables impliquant l'oxygène O9 pour des valeurs de $\tau > 20^\circ$ (Figure 4.14),

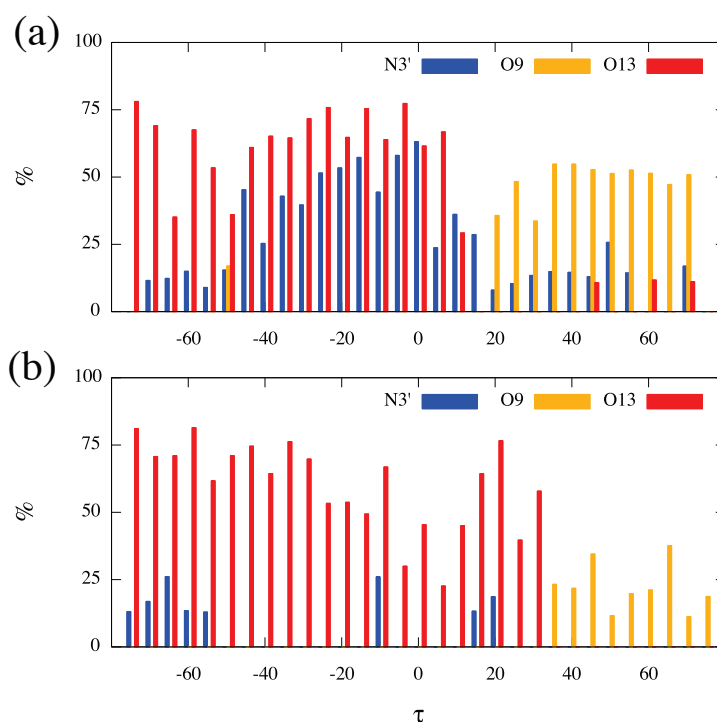


Figure 4.14 – Pourcentage de structure par fenêtre de simulation effectuant une liaison hydrogène impliquant l'azote N3' (en bleu) ou les oxygènes O9 (en jaune) ou O13 (en rouge). Simulations en conditions physiologiques de concentration en sels. **(a)** Le long du chemin EvA. **(b)** Le long du chemin de transition AvE.

corrélée avec une forte stabilisation de la conformation équatoriale du cycle A dans cette zone (Figure 4.13 courbe rouge). Le chemin AvE simulé en conditions physiologiques, contrairement à celui simulé en conditions minimales, présente très peu d'interactions via l'azote N3' et notamment aucune de type pont salin (données non représentées). Cette absence d'interactions permettra au sucre de se réorienter (illustré par la Figure 4.15 représentant les variations des angles τ_1 et τ_2 qui sont proches pour les deux chemins) et ainsi de former des liaisons hydrogène via l'atome O9 lorsque τ sera supérieur à 30° . Les liaisons intermoléculaires impliquant l'oxygène 9 seront toutefois d'une stabilité moindre par rapport au chemin EvA (comme le montre la Figure 4.14), résultant en une plus faible stabilisation de la conformation équatoriale (Figure 4.13). Au niveau structurel, ceci se traduira par une meilleure superposition des structures issues des deux chemins pour des valeurs de τ à proximité du minimum équatorial (voir Figure 4.16).

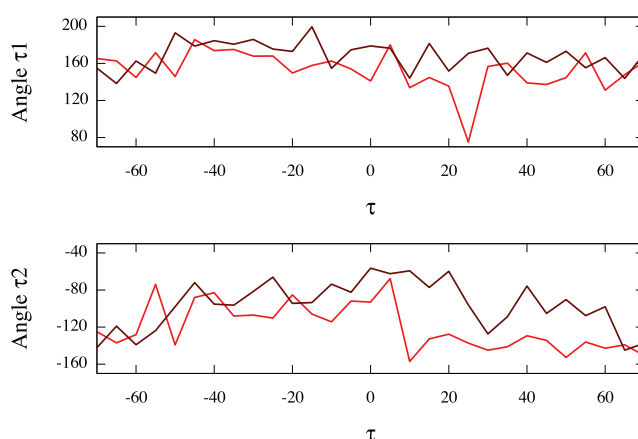


Figure 4.15 – Evolution en degré des dièdres τ_1 et τ_2 issus de structures représentatives de chaque fenêtre d'échantillonnage, le long des chemins de transitions EvA (en rouge) et AvE (en marron) simulés en conditions physiologiques.

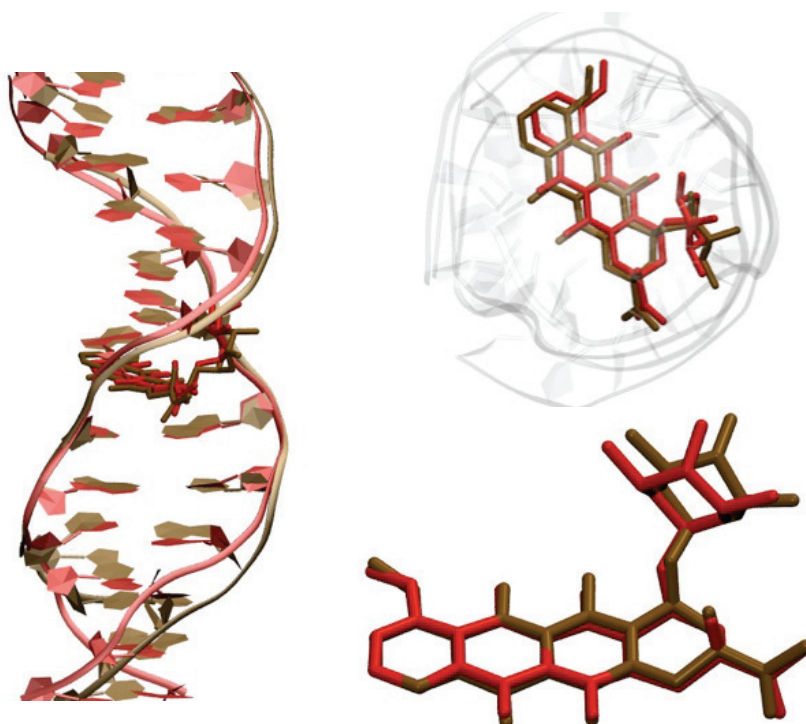


Figure 4.16 – Représentation de structures issues de simulation en conditions physiologiques le long des chemins EvA (en rouge) et AvE (en marron), pour des valeurs du dièdre τ proche du minimum Eq. Les structures sont superposées par rapport aux bases du site d'intercalation, vu du petit sillon (gauche) et vu du haut de la séquence (en haut à droite). En bas à droite, représentation de la superposition des ligands. Ces structures montrent que le ligand se place de façon similaire dans la poche d'intercalation sur les deux chemins.

4.4.2.3 Chemins de transitions et effets de la concentration ionique

Indépendamment des conditions ioniques dans lesquelles nous avons simulé notre système, il apparaît une différence importante au niveau énergétique et structural en fonction du sens de simulation. Le premier enseignement que l'on peut tirer de telles différences est le manque d'échantillonnage. Il s'avère donc très difficile d'échantillonner l'ensemble de l'espace conformationnel et ce, malgré 4 ns de simulations par fenêtre.

Cependant, ces différences que l'on ne retrouve pas lorsque la DNM est simulée librement en milieu aqueux, nous apportent de nombreuses informations sur l'importance des interactions entre les deux partenaires ainsi que sur l'influence de la concentration ionique du milieu.

D'un point de vue énergétique, les chemins simulés en conditions minimales présentent une grande différence d'énergie en fonction du sens de construction du chemin jusqu'à atteindre 9 kcal mol⁻¹ de différence pour les deux conformations équatoriales. En conditions physiologiques, cette différence n'est que de 4 kcal mol⁻¹ alors que le temps d'échantillonnage est le même dans les deux cas. Ce résultat semble montrer une convergence plus rapide des profils d'énergie libre en conditions physiologiques. Ces différences énergétiques trouveront leurs origines au niveau des structures de notre système.

En effet, comme nous l'avons vu précédemment la formation de liaisons hydrogène stables entre l'hydroxyle du ligand lié en C9 et l'ADN est un facteur déterminant de la stabilisation de la conformation équatoriale du cycle A lorsque la DNM est intercalée. Cette interaction s'avère dépendante de la disponibilité de cet hydroxyle, ce dernier pouvant également effectuer une interaction intramoléculaire avec l'oxygène O7 reliant le sucre au cycle A. De là, nous pouvons mettre en avant le rôle important de l'ammonium. En effet, les résultats des simulations en conditions minimales nous montrent que la formation d'un pont salin stable entre l'ammonium et l'ADN sur le chemin AvE a une influence sur la position du sucre dans l'espace permettant la formation de liaisons entre l'hydroxyle en C9 et l'oxygène en C7. Ceci est appuyé par les résultats observés en conditions physiolo-

giques, où pour le chemin AvE, l'absence du pont salin couplé à une réorientation du sucre se traduit par l'apparition de liaison hydrogène entre l'OH9 et l'ADN.

Si l'on compare les interactions intermoléculaires en fonction des conditions de simulation, la formation d'un pont salin stable nous apparaît clairement dépendante de la concentration ionique. En effet, au cours des simulations en conditions minimales, ce pont salin apparaît de façon très stable notamment le long du chemin AvE, tandis que sur le même chemin en conditions physiologiques, il est très peu stable voire inexistant. Ce phénomène s'explique par le fait qu'en milieu physiologique, il y aura une plus grande concentration de cations dans le milieu, notamment au voisinage de l'ADN, pouvant de ce fait interagir avec le squelette fortement électronégatif de l'ADN.

4.4.3 DNM liée au petit sillon de l'ADN

Pour les transitions du dièdre τ lorsque la DNM est liée au petit sillon de l'ADN, la conformation de départ du cycle A que nous avons utilisée est initialement de type Eq, conformation conservée par la DNM à l'issue des simulations de relaxation et d'équilibration des systèmes. Les transitions du cycle A ont été caractérisées dans les mêmes conditions de simulations que précédemment.

4.4.3.1 Transitions du dièdre τ

En conditions minimales, l'angle τ montre une stabilisation de la position axiale indépendamment du chemin d'échantillonnage. Les profils d'énergie libre issus de la transition de l'angle τ (Figure 4.17 en haut à gauche), nous montrent une différence d'énergie de 1 kcal mol⁻¹ en faveur de la conformation axiale pour le chemin EvA (courbe bleue foncée), différence qui sera d'environ 2 kcal mol⁻¹ pour le chemin AvE (courbe bleu claire). Dans les deux cas, les minima sont séparés par une barrière de ~5 kcal mol⁻¹ par rapport à la conformation axiale. Contrairement aux simulations avec la DNM intercalée, les deux chemins montrent des profils

d'énergie libre bien plus convergés ; les conformations équatoriales issues des deux chemins présentant une différence d'énergie $< 2 \text{ kcal mol}^{-1}$.

Les simulations en conditions physiologiques (Figure 4.17 en haut à droite) se caractérisent quant à elles par une stabilisation de la conformation équatoriale dans le sens de simulation EvA (en rouge), avec une différence d'énergie de 2 kcal mol^{-1} entre les deux minima ; la barrière d'énergie séparant ces deux conformations étant de $\sim 5 \text{ kcal mol}^{-1}$ par rapport au minimum le plus stable. Les deux chemins apparaissent cependant moins convergés, le chemin AvE (courbe marron)

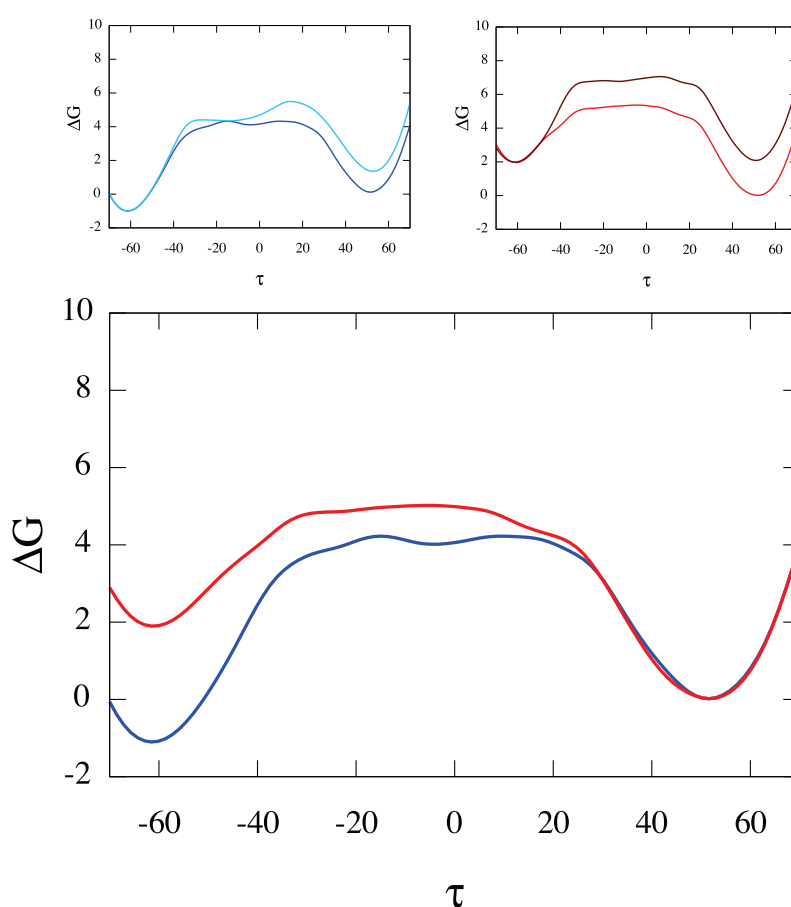


Figure 4.17 – Profils d'énergie libre (kcal mol^{-1}) des transitions conformationnelles de l'angle de torsion τ lorsque la DNM est liée au petit sillon de l'ADN. En haut à gauche : en conditions minimales le long des chemins EvA (bleu foncé) et AvE (bleu clair). En haut à droite : en conditions physiologiques le long des chemins EvA (rouge) et AvE (marron). En bas, comparaison des profils des chemins EvA pour les deux concentrations ioniques (minimale en bleu et physiologique en rouge).

présentant ces deux minima à des énergies équivalentes séparées par une barrière d'énergie de ~ 7 kcal mol⁻¹.

La comparaison des deux profils issus des chemins EvA simulés dans les deux conditions (Figure 4.17 bas) fait clairement apparaître une différence de stabilisation de la conformation Ax du cycle A en fonction de la concentration en sels. Comme nous le verrons dans les paragraphes suivants, ces différences vont s'exprimer par des différences structurales.

4.4.3.2 Influence de la concentration ionique

En considérant uniquement les chemins EvA, la Figure 4.18 nous montre que lorsque la DNM est liée au petit sillon, elle effectue de nombreuses interactions de type pont salin avec le squelette de l'ADN, et ce indépendamment des conditions de simulation (minimales ou physiologiques). Cette différence par rapport aux complexes intercalés sera rendue possible par le fait que lorsque le ligand est lié au petit sillon, seule sa partie conjuguée se trouve dans le petit sillon tandis que le sucre daunosamine est placé à l'extérieur du sillon à proximité du squelette de la double hélice.

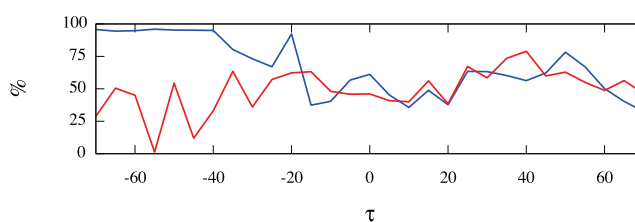


Figure 4.18 – Pourcentage de structures de chaque fenêtre de simulation effectuant un pont salin entre l'ammonium et le squelette de l'ADN le long des chemins EvA simulés en conditions minimales (bleu) et physiologiques (rouge).

Cependant, la Figure 4.18 met également en avant une nette augmentation du nombre de structures effectuant des ponts salins en conditions minimales pour des valeurs de $\tau < -40^\circ$ sur les chemins EvA, zone où les profils d'énergie libre diffèrent le plus. Tout comme dans les simulations avec la DNM intercalée, la présence de

différence dans la stabilité des ponts salins, s'accompagne de différences structurales importantes. Ces différences sont observées non seulement au niveau de la position du sucre illustré par la Figure 4.19 représentant l'évolution des dièdres τ_1 et τ_2 le long des chemins EvA, mais également au niveau de la position de la partie flexible du ligand dans le petit sillon illustrée par la Figure 4.20.

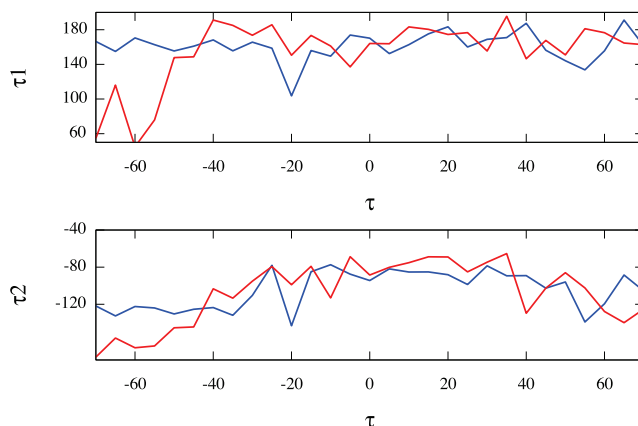


Figure 4.19 – Evolution des valeurs des dièdres τ_1 (en haut) et τ_2 (en bas) exprimées en degrés, en fonction de la valeur de l'angle τ (°), provenant des structures représentatives de chaque fenêtre d'échantillonnage le long des chemins de transitions EvA simulés en conditions minimales (bleu) et physiologiques (rouge).

Lorsque la DNM est liée au petit sillon de l'ADN, certains atomes impliqués dans les interactions interfaciales de type liaisons hydrogène vont naturellement différer de ceux impliqués dans le cas du complexe intercalé (Figure 4.21.c). La position de la partie rigide de la DNM dans le sillon permettra notamment des interactions entre les oxygènes faisant face aux paires de bases tels les oxygènes O11 et O12 de la DNM. Pour toutes les fenêtres d'échantillonnages, l'atome O12 sera d'ailleurs impliqué dans des liaisons intermoléculaires dans plus de 90 % des structures de chaque échantillonnage (ces données ne sont pas représentées sur la Figure 4.21 afin d'en faciliter la lecture).

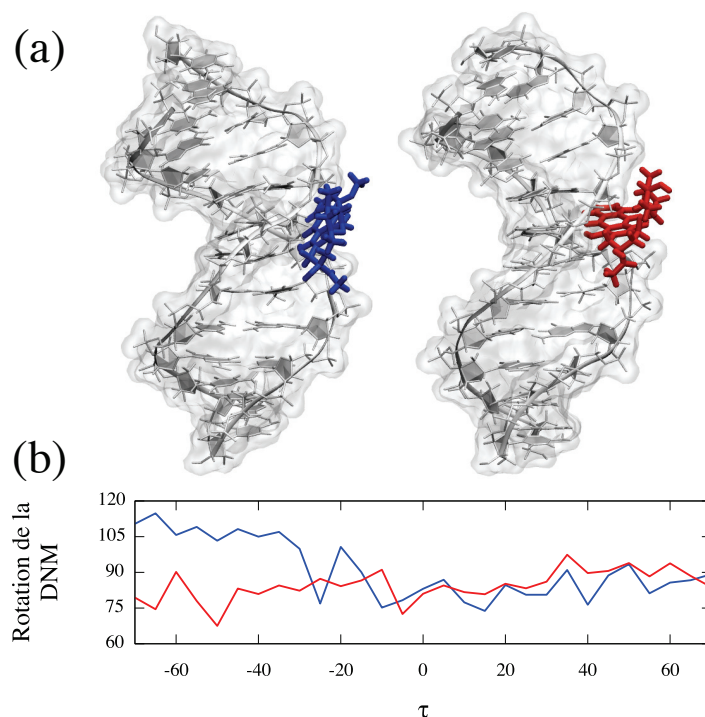


Figure 4.20 – Réorientation de la DNM dans le petit sillon. **(a)** Comparaison des structures proches du minimum axial issues des chemins simulés en conditions minimum (DNM représentée en bleu) et physiologique (DNM représentée en rouge). **(b)** Rotation de la DNM par rapport à la position intercalée pour les structures représentatives de chaque fenêtre d'échantillonnage le long des chemins de transition du dièdre τ ($^\circ$) dans le sens EvA simulées en conditions minimales (en bleu) et physiologiques (en rouge).

Lorsque l'on s'intéresse aux autres atomes impliqués dans de telles liaisons hydrogène, en conditions minimales, la réorientation du ligand dans le petit sillon de l'ADN aux alentours du minimum axial observée sur la Figure 4.20, va permettre la formation de liaisons hydrogène stables via l'hydroxyle OH11 de la DNM ainsi que l'apparition de liaisons hydrogène via l'oxygène O7 (Figure 4.21). De telles interactions étant réduites voire absentes dans la même zone en conditions physiologiques, elles pourront expliquer en partie l'augmentation de stabilité de la conformation axiale en conditions minimales.

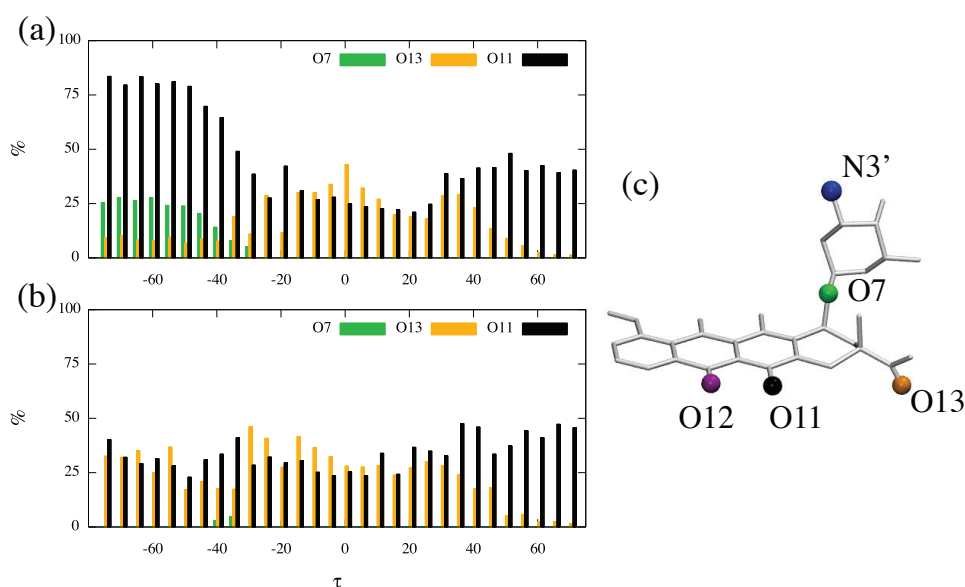


Figure 4.21 – Pourcentage de structures par fenêtre de simulation effectuant une liaison hydrogène impliquant l'azote N3' (en bleu) ou les oxygènes O7 (vert), O13 (jaune) ou O11 (noir) provenant des simulations en conditions (a) minimales et (b) physiologiques de concentration en sels. (c) Représentation des atomes de la DNM impliqués dans des liaisons hydrogène stables lors des simulations.

4.4.3.3 Etudes des paramètres hélicoïdaux de l'ADN le long des chemins de transition

Largeur du petit sillon de l'ADN. Si l'on s'intéresse à la variation de la largeur du petit sillon en fonction de l'évolution du dièdre τ pour les chemins EvA, il nous apparaît qu'indépendamment des conditions de simulation, la conformation équatoriale du cycle A engendre un rétrécissement de la largeur du petit sillon (voir Figure 4.22). En effet, pour des valeurs de τ supérieures à -5° , la majorité des structures présentent des largeurs de sillon au niveau du pas C6pG7 comprises entre 3 Å et 6 Å, valeurs inférieures à celles attendues pour un ADN-B.

Cependant, les profils d'évolution des largeurs de petits sillons (au niveau du site d'intercalation C6pG7), diffèrent en fonction de la concentration. En effet, en conditions minimales, la largeur du petit sillon apparaît clairement couplée à la

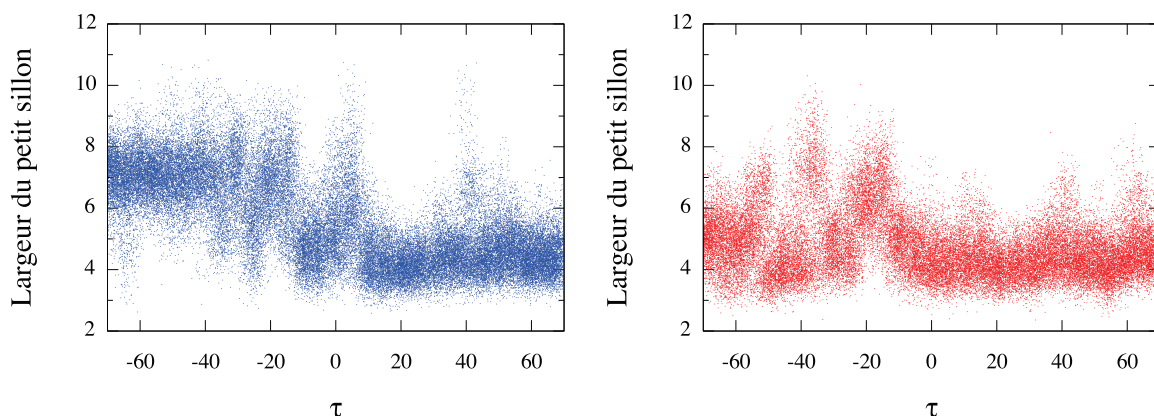


Figure 4.22 – Evolution de la largeur du petit sillon de l'ADN (Å) en fonction de la valeur de τ (°) pour les structures issues des chemins de transition EvA lorsque la DNM est liée au petit sillon et simulée en conditions minimales (à gauche bleu) et physiologiques (à droite rouge).

conformation du cycle A sur le chemin EvA, comme illustré par la Figure 4.22 (gauche) où l'on peut observer une forte augmentation de la largeur du sillon (entre 6,5 et 8 Å) pour des valeurs de $\tau < -30^\circ$. Ce phénomène est fortement atténué en conditions physiologiques (Figure 4.22 droite) ; les structures issues de ces simulations présentant majoritairement des largeurs de petit sillon entre 3 et 6 Å tout au long du chemin de transition.

Cette influence de la concentration ionique sur les couplages entre la largeur du petit sillon de l'ADN et la conformation du cycle A est indépendante du sens de simulation puisqu'elle se retrouve également pour les simulations dans le sens AvE (résultats non représentés).

Influence sur le Roll. La présence de la DNM dans le petit sillon aura une influence sur les paramètres de l'ADN, et notamment sur le « Roll ». Le *roll* correspond à l'angle créé par deux paires de bases dans la direction du petit sillon ; dans un ADN-B canonique ainsi que dans nos simulations de la séquence d'ADN libre (Figure 4.4), sa valeur est proche de 0.

Indépendamment de la concentration en sels et du sens de simulation, lorsque le cycle A de la DNM se trouve en conformation E_q, le *roll* du pas C6pG7 montre une valeur d'angle négative correspondant à un *roll* ouvert vers le grand sillon (Figure 4.23 – le *roll* évoluant de façon similaire dans le sens EvA et AvE, seuls les chemins EvA sont représentés). Le long de nos chemins de transition, les valeurs de *roll* augmentent en même temps que les valeurs de τ diminuent, pour finalement obtenir un *roll* positif ouvert vers le petit sillon lorsque le cycle A de la DNM se trouve en conformation A_x.

Selon l'étude menée par Mukherjee et ses collaborateurs en 2008 [227], l'augmentation du *roll* est l'une des premières étapes conduisant à l'intercalation de la DNM dans son site d'intercalation.

L'influence de la conformation du cycle A de la DNM sur la valeur du *roll* pourrait être le reflet d'un mécanisme important dans le processus d'intercalation. En effet, lorsque la DNM se trouve dans le petit sillon, la transition de son cycle A vers une conformation de type A_x engendrant une augmentation du *roll*, pourra être assimilée à une première étape en vue de l'intercalation de la molécule.

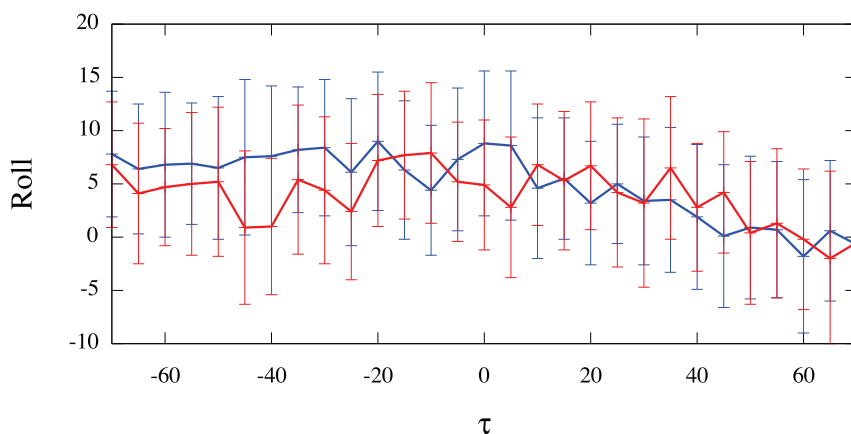


Figure 4.23 - Evolution du *roll* (°) pour le pas C6pG7 des structures représentatives de chaque fenêtre d'échantillonnage sur les chemins de transition EvA lorsque la DNM est liée au petit sillon en conditions minimales (en bleu) et physiologiques (en rouge)

4.5 Discussion et conclusions

L'étude des transitions du cycle A lorsque la DNM se trouve libre ou complexée à l'ADN (que ce soit intercalée ou liée au petit sillon), nous permet de mettre en avant une forte influence des interactions intermoléculaires, notamment vis à vis de l'équilibre des conformations du cycle A.

Libre dans le milieu aqueux, le cycle A de la DNM montre un équilibre légèrement en faveur d'une conformation axiale du cycle A.

Intercalé, en plus d'une forte dépendance au sens de construction des chemins de simulations (conséquence directe de la difficulté d'échantillonnage de tout l'espace conformationnel), les conformations du cycle A seront fortement influencées par la présence de liaisons intermoléculaires. Notamment, la présence de liaisons hydrogène interfaciales stables aura tendance à favoriser la conformation équatoriale du cycle A au détriment de la conformation axiale.

Lié au petit sillon, le cycle A du ligand présente une stabilité conformationnelle qui s'avère être dépendante de la concentration ionique, avec une conformation Eq favorisée en conditions physiologiques et Ax en conditions minimales.

L'ensemble de ces résultats nous ont également permis de pointer du doigt l'importance non négligeable de la concentration ionique. En effet, alors que la concentration saline ne semble avoir aucune influence notable sur chacun des partenaires isolés, les simulations des complexes en concentrations minimales nous montrent une tendance à effectuer des interactions très stables de type pont salin à l'interface. Ce type d'interactions étant réduit en conditions physiologiques. Ces différences pourront s'expliquer par le fait que l'ADN, de par la présence de son squelette phosphate, est très négativement chargé. En conséquence, la présence d'une charge positive portée par le groupement ammonium de la DNM permettra de créer de fortes interactions entre ce dernier et la double hélice. Ces interactions seront encore renforcées par le faible nombre d'ions Na^+ rencontrés en conditions minimales. Au contraire, en milieu plus riche en sels, la présence d'un plus grand nombre d'ions Na^+ à proximité du squelette permettra de compenser

l'environnement fortement électronégatif créé par les phosphates et de réduire les interactions impliquant l'ammonium.

Au cours de nos simulations, l'augmentation de stabilité des ponts salins s'accompagne de réorientation du ligand. Ces réorientations ont non seulement montré des conséquences dramatiques vis à vis des interactions entre les deux partenaires, mais également une influence non négligeable sur la conformation de l'ADN (comme nous le montrent les variations de largeur du petit sillon de ce dernier lorsque la DNM s'y trouve liée).

En conséquence de quoi, afin de réduire la surestimation apparente des interactions interpartenaires dues au manque de cations dans le système, tous les travaux que nous avons effectués ultérieurement ont été réalisés en condition physiologique.

Finalement, les transitions du cycle A lorsque la DNM est liée au petit sillon nous ont également permis, de par la corrélation entre la conformation du cycle A et du *roll*, de mettre en lumière la potentielle implication de transition du cycle A dans l'initiation de l'intercalation. L'observation de ce phénomène lorsque le ligand est lié au petit sillon, nous permet d'appuyer l'existence de cet état intermédiaire, qui reste hypothétique à l'heure actuelle.

Tous ces éléments nous permettant d'envisager le processus d'intercalation de la DNM sous un nouveau jour.

CHAPITRE 5

Etude du chemin de désintercalation

5.1 Introduction

L'étude menée par Mukherjee en 2008 en collaboration avec Lavery, Bagchi et Hynes [227], a permis d'obtenir le profil d'énergie libre de désintercalation de la DNM entre deux paires de bases de l'ADN avec des énergies proches de celles obtenues par des méthodes expérimentales. Ces travaux de dynamique moléculaire ont mis en évidence un mécanisme d'intercalation menant la DNM d'un état libre dans le milieu aqueux à son site d'intercalation en passant par un état stable où la DNM se trouve liée au petit sillon. Cet état, entraînant peu de déformations au niveau de l'ADN, semble correspondre à l'étape de liaison « extérieure » proposée dans le modèle de Chaires [219]. Le passage de cet état lié au petit sillon vers le site d'intercalation nécessite un apport d'énergie de ~ 12 kcal mol⁻¹. Le processus décrit s'accompagne de distorsions de la double hélice impliquant des modifications des paramètres de *roll* et de *rise* ainsi qu'une modification des motifs de liaisons hydrogène impliquant les molécules d'eau.

Cependant, bien qu'elles permettent d'obtenir les énergies libres de ces deux états, les contraintes utilisées au cours de cette étude pour dissocier les complexes ne permettent pas de fournir des détails quant au chemin menant directement la DNM de l'état intermédiaire liée au petit sillon vers l'état intercalé.

La construction et la caractérisation de ce chemin sont donc apparues comme la suite logique de ces travaux vers la compréhension du processus d'intercalation dans son ensemble.

Nous avons donc créé le chemin d'énergie libre reliant ces deux points par la technique d'*umbrella sampling* couplée à un potentiel de contrainte, permettant de biaiser le système vers l'état final que nous désirons. Pour cela, nous avons utilisé une méthode développée dans notre laboratoire.

Dans le cadre de travaux visant à appréhender les mécanismes guidant les interactions entre la protéine SRY et l'ADN, Benjamin Bouvier en collaboration avec Richard Lavery [319], a mis au point une contrainte de séparation imposant une distance minimale donnée entre deux partenaires (en tenant compte uniquement

des paires d'atomes non hydrogène). L'utilisation de cette contrainte nous a permis de séparer la daunomycine de l'ADN sans avoir à spécifier des coordonnées spatiales guidant le chemin de séparation et ainsi permettre à la DNM de pouvoir éventuellement se réorienter dans l'espace et ainsi se lier au petit sillon.

En plus de la contrainte de séparation, nos travaux ont également permis d'intégrer des modifications par rapport aux simulations publiées par Mukherjee *et al.* Les dynamiques effectuées au cours de ces travaux initiaux ont toutes été réalisées avec le champ de force AMBER94. Comme nous l'avons vu précédemment, ce champ de force pourra être à l'origine de stabilisation de conformations des angles de torsion α/γ de l'ADN [73], pouvant avoir des effets dramatiques sur la double hélice lors de longues simulations. Pour nos simulations, nous avons donc utilisé le champ de AMBER parm99 avec la modification bsc0 apportant une correction à la description du squelette de l'ADN [75].

Egalement, comme nous l'avons vu dans la partie précédente, en conditions minimales de concentration en sels il existe un couplage important entre la conformation du cycle A et la largeur du petit sillon lorsque le ligand y est lié. De ce fait, nos simulations ont été réalisées dans des conditions physiologiques, alors que les simulations de la publication initiale ont été réalisées en conditions minimales [227].

En complément à cette étude, des travaux d'Arnab Mukherjee ont également permis de caractériser l'énergie libre du système par métadynamique. Le chemin réactionnel obtenu (qui fera l'objet d'une brève discussion ci-dessous), s'avère être cohérent avec nos résultats d'*umbrella sampling*.

L'ensemble de ces travaux a fait l'objet d'une publication commune sous le titre « *A multistep intercalation mechanism : molecular dynamics and free energy studies of the formation of a ADN-daunomycin complex* » parue en 2012 dans *Journal of the American Chemical Society* [307] (jointe en Annexe B).

5.2 Description de la contrainte de séparation

Pour simuler la sortie de la DNM de son site d'intercalation nous nous sommes appuyés sur une contrainte de séparation initialement développée et utilisée avec succès pour séparer des complexes ADN-protéines [319,320].

Cette contrainte consiste à imposer une distance minimale entre deux groupes d'atomes disjoints, autre que les atomes d'hydrogène (entre les atomes interfaciaux de deux partenaires d'un complexe par exemple).

À un pas de temps donné, la contrainte agira sur toutes les paires d'atomes (avec un atome dans chaque groupe) selon le potentiel de biais total E , qui correspond à la somme des potentiels individuels de chaque paire d'atomes pour laquelle $d_i < d_{min}$ à ce même pas de temps.

$$E = \sum E_i \quad (5.67)$$

avec :

$$\begin{cases} E_i = k(d_i - d_{min})^2 & d_i < d_{min} \\ E_i = 0 & d_i \geq d_{min} \end{cases} \quad (5.68)$$

où i représente une paire d'atomes, k la constante de force, d_i la distance euclidienne séparant les deux atomes de la paire et d_{min} la distance minimale de séparation choisie.

Les forces correspondantes agissant sur les atomes A_i et B_i d'une paire d'atomes, pourront s'exprimer sous la forme suivante pour $d_i < d_{min}$:

$$\begin{cases} \mathbf{f}_i^A = -2k(d_i - d_{min}) \cdot \mathbf{u}_i \\ \mathbf{f}_i^B = 2k(d_i - d_{min}) \cdot \mathbf{u}_i \end{cases} \quad (5.69)$$

avec,

$$\mathbf{u}_i = \frac{\mathbf{A}_i \mathbf{B}_i}{\|\mathbf{A}_i \mathbf{B}_i\|} \quad (5.70)$$

Afin d'éviter que les deux groupes d'atomes ne s'éloignent l'un de l'autre de façon non contrôlée lorsque $d_i \geq d_{min}$ pour toutes les paires d'atomes (notamment aux grandes distances de séparation où ne subsistent que de faibles interactions liantes), une pénalité quadratique similaire est appliquée sur la paire d'atomes les plus proches uniquement. Le potentiel sera alors de la forme :

$$E = k(d_c - d_{min})^2 \quad (5.71)$$

avec $d_c = \min(d_i)$.

5.3 Chemin de passage de la DNM de son site d'intercalation au petit sillon de l'ADN par umbrella sampling

Nous avons construit le chemin menant la DNM de sa position intercalée (IC) entre deux paires de bases de l'ADN vers un état lié au petit sillon (MG1) à partir de deux chemins différents. Un menant la DNM hors de son site d'intercalation, tandis que le second nous a permis de séparer les deux partenaires lorsque la DNM est liée au petit sillon.

5.3.1 Séparation à partir de la position IC

La première série de simulations consiste à retirer le ligand de son site d'intercalation. Pour ce faire nous avons défini les atomes des bases de l'ADN entre lesquelles est intercalée la DNM (exception faite des hydrogènes) en tant que groupe d'atomes 1, et les atomes de la partie rigide du cycle A (*i.e.* les atomes C7, C6a, C10a et C10) de la DNM en tant que groupe d'atomes 2. Le choix des atomes du second groupe est guidé par la nécessité d'inciter le ligand à sortir vers le petit sillon, tout en le laissant libre de se réorienter une fois séparé de la cavité d'intercalation (Figure 5.1).

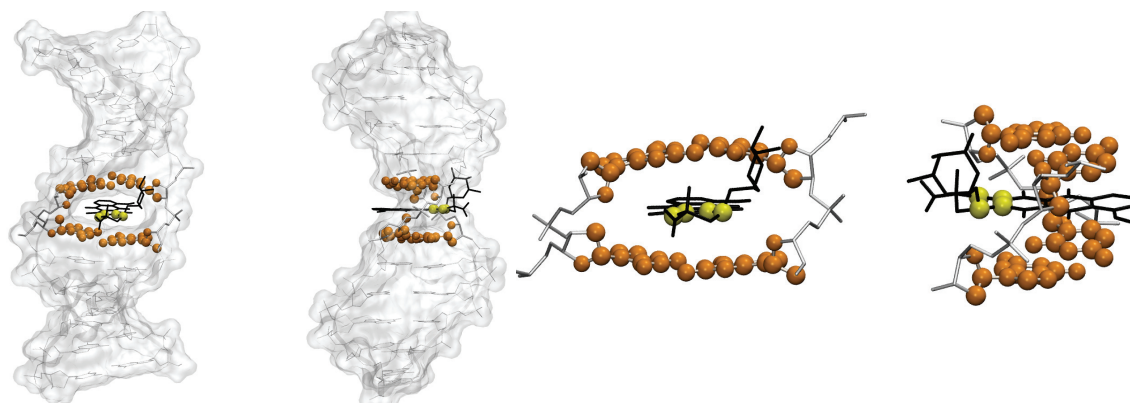


Figure 5.1 – Représentation des atomes entre lesquels la contrainte de distance minimale est appliquée pour le chemin de séparation IC. Le groupe d'atomes impliquant les atomes des bases du site d'intercalation est représenté en orange ; le groupe d'atomes impliquant des atomes du ligand est représenté en jaune.

A partir de la structure de la DNM intercalée utilisée pour simuler la transition du cycle A dans la partie précédente (voir partie 4.1), nous avons réalisé une série de 64 fenêtres de simulation en imposant des distances minimales de séparation allant de 3,2 Å à 12,5 Å par pas de 0,15 Å, avec des constantes de forces de 50 kcal mol⁻¹ Å⁻². Le pas de 0.15 Å choisi nous permet d'avoir un bon recouvrement des échantillonnages (voir Annexe C). Le point de départ d'une fenêtre n ($n > 1$) étant la structure finale du point $n-1$ après 1 ns ou 2 ns de simulation en fonction du degré de liberté du ligand. Les échantillonnages de chaque fenêtre ont ensuite été augmentés à 4 ns ou 8 ns pour vérifier la convergence des profils d'énergie libre.

De nombreuses optimisations ont été nécessaires à l'obtention du chemin de séparation final. En effet, les premières tentatives de séparation des partenaires ont été réalisées avec des paramètres de simulation différents de ceux utilisés finalement (avec par exemple des temps de simulation plus courts, des pas entre chaque fenêtre d'échantillonnage plus grands, ou encore des constantes de force différentes). Dans certains cas, ces tentatives ont notamment abouti à des déformations de l'ADN, que nous avons attribuées à un échantillonnage et une relaxation insuffisantes de certaines fenêtres de simulation avant de passer à la suivante.

De même, le premier chemin que nous avons construit, nous a permis de sortir la DNM de son site d'intercalation ; cependant un trop faible temps de simulation entre les différentes fenêtres ne nous a jamais permis d'atteindre des états conformationnels qui, nous le verrons par la suite, jouent un rôle essentiel dans le processus d'intercalation.

Par conséquent, nous avons adapté le temps d'échantillonnage ainsi que les autres paramètres de simulation comme décrit précédemment pour obtenir le chemin d'énergie libre présenté dans la partie suivante.

5.3.1.1 Profil d'énergie libre

La Figure 5.2 représente le profil d'énergie libre du chemin de sortie de la DNM de son site d'intercalation. Le minimum d'énergie est obtenu pour une valeur de contrainte de 3,2 Å correspondant à la position optimale de la DNM intercalée dans le petit sillon. A partir de ce point, plusieurs phases semblent caractériser la séparation des deux partenaires. Une première phase de perte d'énergie linéaire entre 3,2 Å et 5,8 Å. Cette perte d'énergie est le reflet de la sortie de la DNM de son site d'intercalation, où son cycle anthraquinone est inséré de façon parallèle entre deux paires de bases, provoquant la rupture des interactions de type π -stacking entre les bases et le cycle conjugué de la DNM. A partir de 5,8 Å (la structure correspondante est représentée en Figure 5.3.b), la DNM est suffisamment désintercalée pour pouvoir entamer une rotation par rapport à son axe de départ. Ce phénomène se traduit par un maximum local d'énergie de 9 kcal mol⁻¹ par rapport à l'état intercalé, suivi par un état intermédiaire métastable. Cet état intermédiaire se caractérise par la présence d'un puits d'énergie entre 5,8 Å et 7,4 Å centré sur un minimum local de 7,7 kcal mol⁻¹ pour une valeur de contrainte de 6,8 Å. L'augmentation supplémentaire de la distance interpartenaire entre 7,4 Å et 12,5 Å s'accompagne d'une lente perte d'énergie d'interaction, pour atteindre un maximum de 13 kcal mol⁻¹. Cette énergie libre d'intercalation de -13 kcal mol⁻¹ est cependant supérieure (bien que proche) des estimations expérimentales (~-9,4 kcal mol⁻¹) [217], ainsi que de celle obtenue par Mukherjee *et al.*, (-12,3 kcal mol⁻¹).

Cette différence d'énergie libre s'explique par la difficulté d'obtention d'un échantillonnage complet de l'espace lorsque la DNM commence à se séparer de l'ADN.

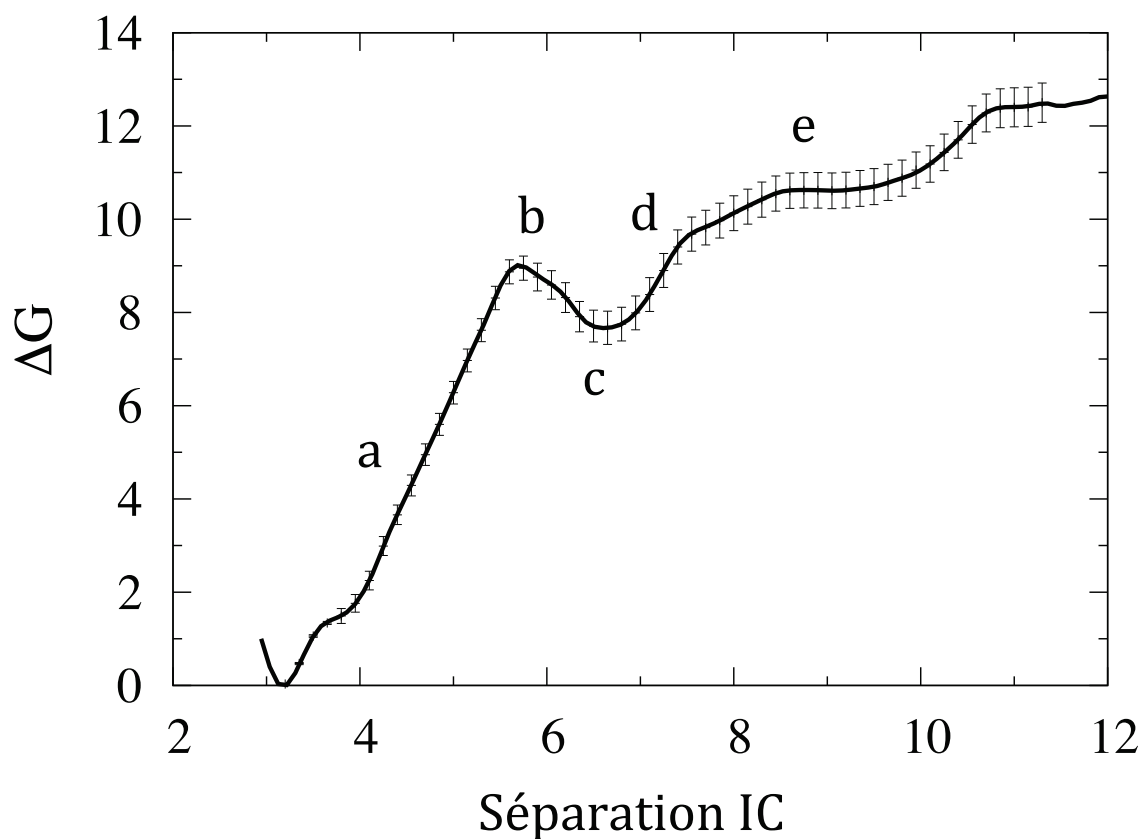


Figure 5.2 – Profil d'énergie libre (en kcal mol⁻¹) du chemin de sortie du petit sillon obtenu à partir des simulations d'*umbrella sampling*. Les lettres a – e correspondent aux différents groupes structuraux décrits dans la partie 5.3.1.2 et la Figure 5.3. Les barres d'erreur correspondent à deux fois l'écart-type de l'énergie libre le long du chemin, calculée selon Zhu et Hummer [305].

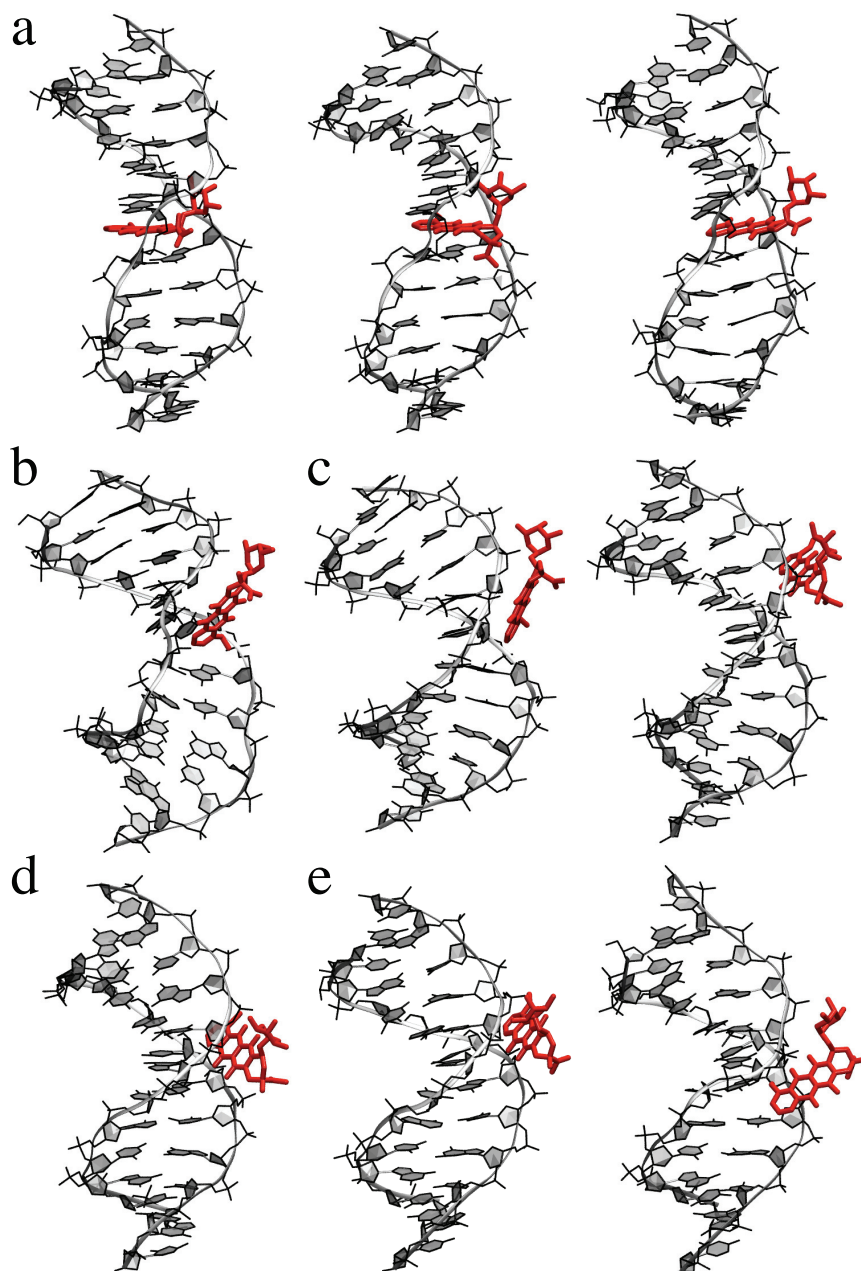


Figure 5.3 – Structures représentatives issues des cinq groupes structuraux identifiés le long du chemin de sortie à partir de la position IC. Les lettres correspondent aux phases d'où sont issues les structures suivantes.

5.3.1.2 Description du chemin de séparation

La présence de plusieurs phases distinctes lors du processus de désintercalation est appuyée par des calculs de RMSD-2D illustré par la Figure 5.4.

Cette RMSD-2D des structures représentatives de chaque fenêtre d'échantillonnage le long du chemin réactionnel, permet de mettre en évidence cinq groupes structuraux. Avec l'appui du profil d'énergie libre (Figure 5.2) et des structures représentatives (Figure 5.3) nous pouvons classer ces cinq groupes de la façon suivante : (a) DNM intercalée ou partiellement intercalée (échantillonnages entre 3,2 Å et 5,8 Å) ; (b) premier état de transition (entre 5,8 Å et 6,2 Å) ; (c) intermédiaire métastable (entre 6,2 Å et 6,5 Å) ; (d) second état de transition (entre 6,5 et 7,4 Å) ; (e) séparation (à partir de 7,4 Å et au-delà).

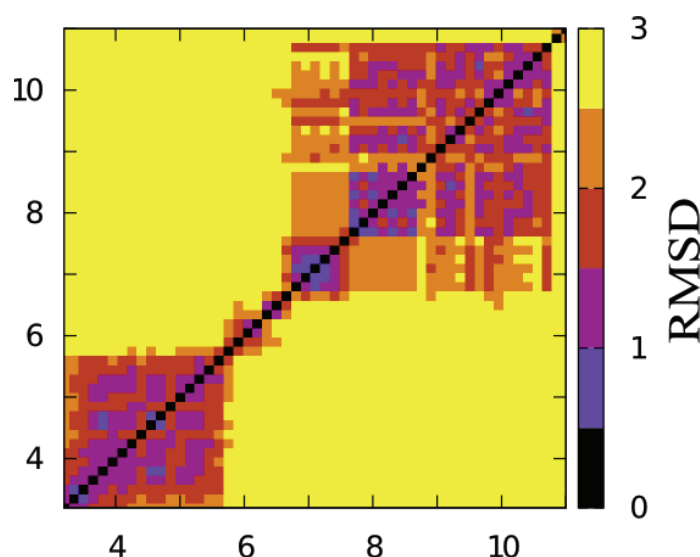


Figure 5.4 – RMSD-2D entre les structures représentatives des échantillonnages le long du chemin de séparation à partir de la DNM intercalée. Les mesures de RMSD sont réalisées en Å en considérant les atomes lourds de la DNM et des six paires de bases du centre de la séquence d'ADN (2 paires de bases impliquées dans l'intercalation et 2 paires de bases au-dessus et en dessous).

Au cours de la phase (a), mis à part des transitions du cycle A de la DNM qui passe d'une conformation équatoriale vers une conformation axiale, aucun événement structurel majeur n'est à noter concernant l'ADN. Lorsque le système atteint le

premier état de transition (b), on peut observer une modification des paramètres de *roll* (augmentation nette jusqu'à $\sim 60^\circ$) et de *rise* (qui passe de ~ 8 Å à ~ 6 Å) au niveau du pas C6pG7 (Figure 5.1). Ces changements de l'ADN déjà observés par Mukherjee *et al.* [227], sont synonymes de fermeture du site d'intercalation vers le grand sillon. Ils s'accompagnent d'une rotation du cycle anthraquinone de la DNM autour de l'axe O5-O12 afin de permettre au ligand de garder les interactions entre son cycle D et les bases de l'ADN. Ce mouvement est continu jusqu'à atteindre l'intermédiaire métastable (c). Le ligand finit par tourner entièrement pour s'aligner avec le petit sillon, accompagné par la fermeture complète de la cavité d'intercalation (illustrée par un retour des paramètres de *rise*, *roll* et de la largeur du petit sillon pour le pas C6pG7 à des valeurs proche d'un ADN-B canonique, Figure 5.5). Il est important de noter que la présence d'un intermédiaire métastable entre deux états de transition n'a pas été mis en évidence dans [227].

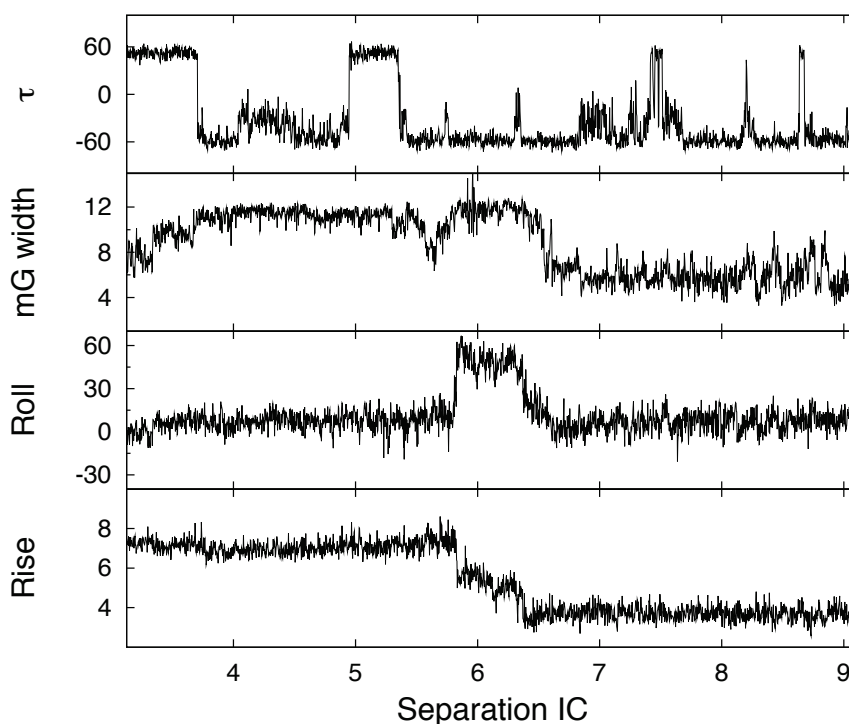


Figure 5.5 – Evolution de l'angle τ ($^\circ$) de la DNM et des paramètres de *rise* (Å), *roll* ($^\circ$) et de la largeur du petit sillon (Å) pour le pas C6pG7 de l'ADN le long du chemin de séparation à partir de la DNM intercalée.

5.3.2 Séparation à partir de la position MG1

La seconde série d'échantillonnage parapluie consiste à séparer les deux partenaires lorsque la DNM se trouve liée au petit sillon.

Pour cela, à partir de la structure utilisée pour caractériser la transition du cycle A lorsque la DNM est liée au petit sillon (voir paragraphe 4.1), nous avons imposé une distance minimale entre tous les atomes lourds des bases de l'ADN d'une part, et tous les atomes lourds de la DNM d'autre part. La distance initiale entre le ligand et l'ADN étant de 2,8 Å, nous avons réalisé 31 fenêtres d'échantillonnage avec des valeurs de contrainte allant de 2.6 Å pour la première fenêtre de simulation, à 7.1 Å par pas de 0.15 Å (voir Annexe C pour le recouvrement des échantillonnages). Pour chaque fenêtre n , nous avons réalisé un échantillonnage initial de 2 ns dont la structure finale nous a servi de point de départ pour la fenêtre $n+1$, chaque simulation étant étendue à 6 ns par la suite.

5.3.2.1 Profil d'énergie libre du chemin de séparation MG1

Le PMF du chemin de séparation MG1 illustré par la Figure 5.7 présente un minimum d'énergie localisé pour une valeur de contrainte de 2,8 Å. Autour de ce minimum, le profil est quasi-quadratique avec une perte d'énergie d'interaction importante (de l'ordre de 6 kcal mol⁻¹) pour des valeurs de contrainte situées entre 2,8 Å et 3,5 Å. Lorsque les partenaires sont séparés de plus de 3,5 Å, la réduction des interactions entre la DNM et le petit sillon s'accompagne d'une augmentation d'énergie linéaire. Ce changement observé sur le profil d'énergie libre se retrouve au niveau structural et notamment lorsque l'on s'intéresse à la RMSD-2D des structures représentatives de chaque fenêtre d'échantillonnage illustré par la Figure 5.6.

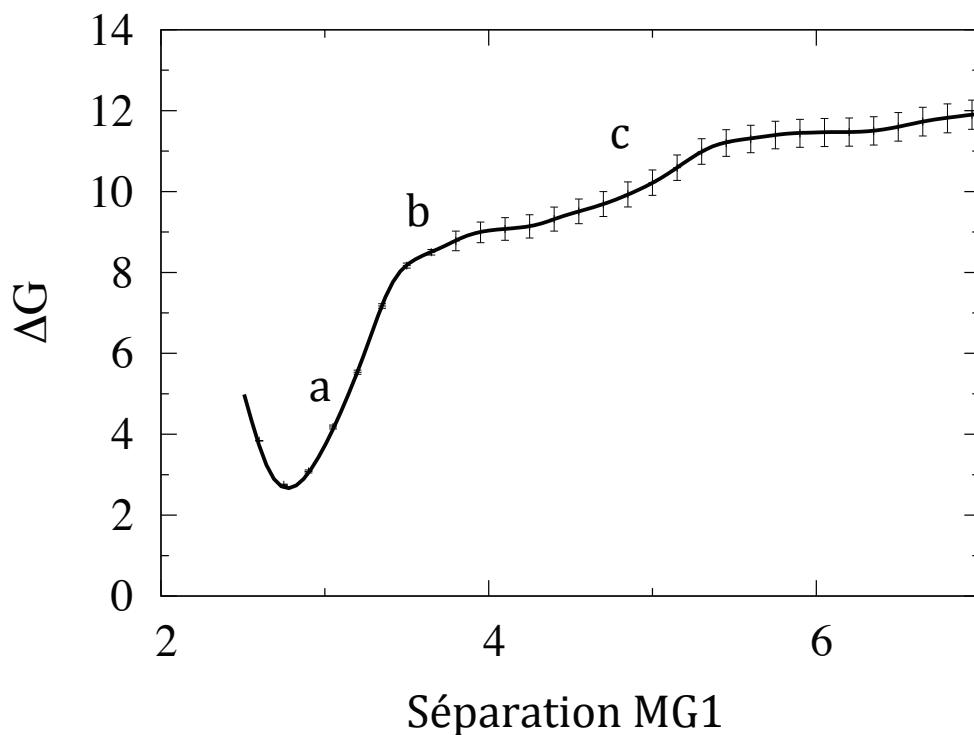


Figure 5.7 – Profil d'énergie libre (en kcal mol⁻¹) du chemin de séparation des deux partenaires lorsque la DNM est liée au petit sillon à partir des simulations d'*umbrella sampling*. Les lettres a – c sont décrites dans la Figure 5.8.

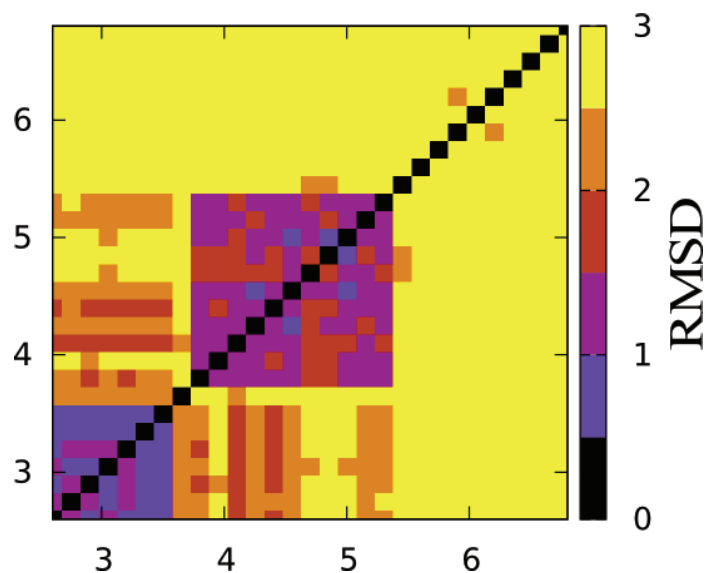


Figure 5.6 – RMSD-2D (Å) entre les structures représentatives des fenêtres d'échantillonnage le long du chemin de séparation à partir de la DNM liée au petit sillon. Mesures réalisées en considérant les atomes lourds de la DNM et des six paires de bases du centre de la séquence d'ADN.

5.3.2.2 Description du chemin de séparation

Si l'on considère ce chemin de séparation d'un point de vue conformationnel, la perte des fortes interactions entre la DNM et l'ADN s'accompagne d'une brusque augmentation de la largeur du petit sillon au niveau du pas d'intercalation (de l'ordre de 2 - 3 Å), ainsi que d'une augmentation de la fluctuation de la largeur de ce sillon (Figure 5.8), qui est quasi nulle lorsque la DNM y est profondément enfouie. Ces changements au niveau de l'ADN s'accompagnent de changements dans le comportement du cycle A de la DNM. En effet, lorsque la distance entre la DNM et l'ADN est faible, celui-ci se trouve dans une conformation équatoriale stable, tandis que par la suite il fluctue entre des conformations axiales et équatoriales.

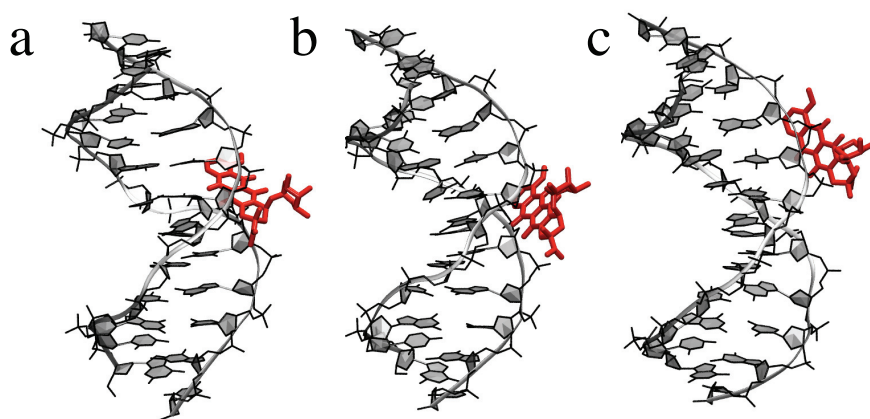


Figure 5.8 – Structures représentatives de zones où la valeur de contrainte est inférieure à 3,5 Å (a) ; égale à 3,5 Å (b) ; supérieure à 3,5 Å (c), le long du chemin de séparation des partenaires à partir de la position MG1.

Par la suite, le ligand ne sort pas du sillon selon un mouvement de translation pure : son éloignement de l'ADN s'accompagne de rotations autour du squelette de l'ADN afin de maintenir le pont salin se formant entre un groupement phosphate anionique du squelette et le groupement amino cationique du sucre daunosamine du ligand.

La Figure 5.9 nous montre que lorsque la DNM se trouve liée au petit sillon, la dissociation du complexe n'a qu'un impact limité sur les paramètres de l'ADN.

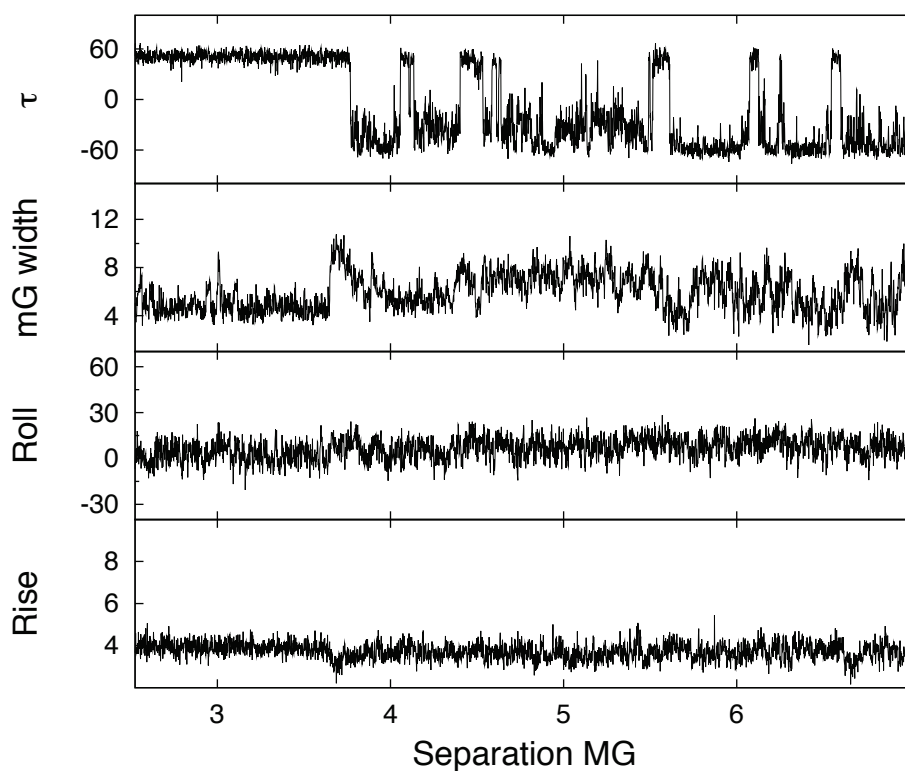


Figure 5.9 – Evolution de l'angle τ (°) de la DNM ainsi que de la largeur du petit sillon (Å) et des paramètres de *roll* (Å) et *rise* (°) et pour le pas C6pG7 de l'ADN le long du chemin de séparation MG1 (Å).

5.3.3 Passage de la DNM de son site d'intercalation au petit sillon de l'ADN

Dans les travaux originels publiés par Mukherjee, la DNM a pu être entièrement séparée de l'ADN, que ce soit à partir de l'état intercalé ou lié au petit sillon. Ceci a permis de caractériser les variations d'énergie libre nécessaires à ces systèmes pour passer d'un état complexé à un état où les interactions entre les deux partenaires sont quasi nulles. De ce fait, les deux chemins de séparations obtenus ont pu être reliés permettant ainsi d'estimer la barrière d'énergie entre l'état MG1 et IC ($\sim 11,9$ kcal mol⁻¹).

Dans nos travaux, la contrainte utilisée pour séparer les deux partenaires n'impose aucune direction de séparation. Dans le cas du chemin de séparation MG1, la contrainte étant appliquée entre tous les atomes des bases de l'ADN et tous les atomes du ligand, une séparation complète des deux molécules est donc envisageable. Pour le chemin de séparation IC, les groupes d'atomes choisis imposent une séparation entre une partie de la DNM et le centre du site d'intercalation (soit à peu près le centre de la structure d'ADN). Dans ce cas, couplée à de fortes interactions entre les deux partenaires, l'augmentation de la distance minimale entre les deux groupes d'atomes permettra de sortir le ligand de son site d'intercalation, mais ne garantira pas une séparation complète de l'ADN et de son ligand. De plus, l'augmentation de la distance interpartenaire augmentant l'entropie rotationnelle des partenaires, l'échantillonnage complet de l'espace sera d'autant plus difficile et coûteux en temps de calcul. De ce fait, il nous est donc impossible de relier les deux chemins de séparation en considérant l'énergie libre nécessaire à la séparation complète des deux partenaires.

Cependant si l'on considère les structures issues des deux ensembles de simulations par RMSD-2D (Figure 5.10.a), les deux chemins s'avèrent présenter des structures très similaires (avec une RMSD < 1 Å) pour des valeurs de contrainte de 7,25 Å sur le chemin IC et 4,1 Å dans le cas du chemin MG1. Les atomes sur lesquels agit la contrainte étant différents, il est donc normal que les chemins puissent présenter des structures identiques pour des valeurs de contrainte très différentes. Cette correspondance de structure est renforcée par la Figure 5.10.b où la superposition des deux structures montre une forte similarité, ainsi que par les valeurs des paramètres de l'ADN (Tableau 5.1) qui sont très proches dans les deux cas.

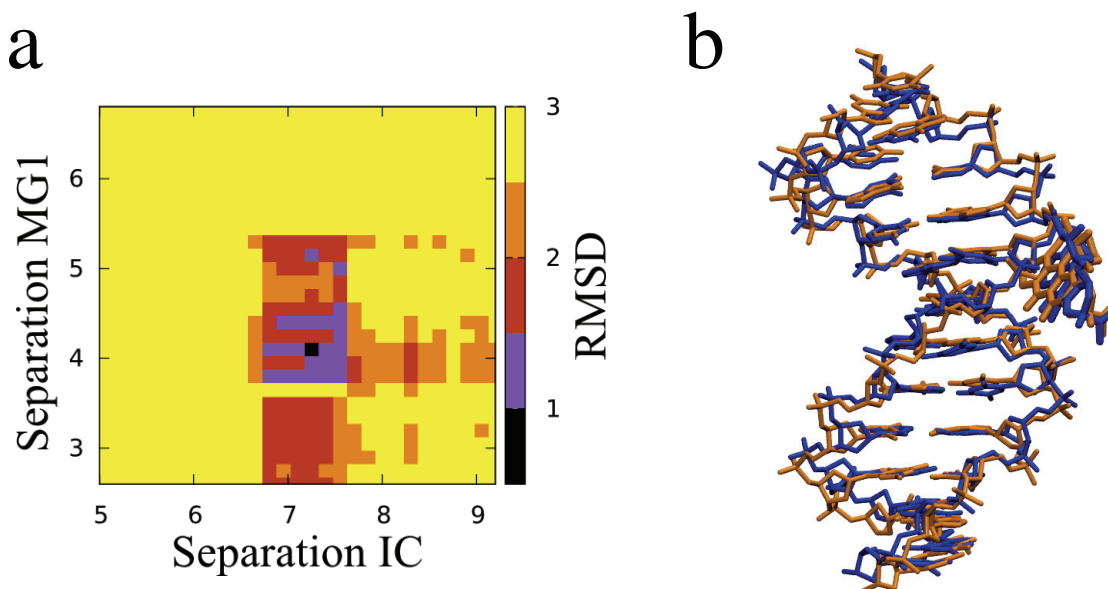


Figure 5.10 – (a) RMSD-2D (Å) entre les structures représentatives des échantillonnages le long du chemin MG1 et du chemin IC. (b) Superposition des structures représentatives des fenêtres à 7,25 Å (chemin IC) en orange et 4,1 Å (chemin MG1) en bleu.

Paramètres	Chemin IC	Chemin MG1
Rise C6pG7 (Å)	3,6	3,9
Roll C6pG7 (°)	2	2
\sum Roll A5-T8 (°)	6	2
Twist C6pG7 (°)	35	38
\sum Twist A5-T8	95	99
Largeur du petit sillon (Å)	6	6
Cycle A	Ax	Ax

Tableau 5.1 – Comparaison des paramètres de l'ADN pour les structures représentatives de la fenêtre de simulation centrée sur 7,25 Å le long du chemin IC et la fenêtre centrée sur 4,1 Å le long du chemin MG1.

La similarité des structures à ce niveau, et le fait que les énergies libres soient déterminées à une constante additive près, nous permettent de faire la jonction entre les deux chemins et ainsi créer un nouveau chemin menant la DNM de son site d'intercalation à un état lié au petit sillon. De même, il nous est possible de réunir les deux profils d'énergie libre pour n'en faire plus qu'un illustré par la Figure 5.11.

Ce nouveau profil d'énergie libre fait apparaître en plus du premier état de transition décrit précédemment, un second état de transition localisé au niveau de la jonction des deux profils. Les deux états de transition présentant des énergies identiques (~ 9 kcal mol⁻¹ par rapport à l'état intercalé). La jonction de deux segments d'énergie libre nous permet également de placer le minimum d'énergie de l'état lié au petit sillon à une énergie libre de 2,7 kcal mol⁻¹ toujours par rapport à l'état intercalé. Les figures représentatives issues de ce nouveau chemin ainsi que les paramètres leur correspondant, seront illustrés par la Figure 5.14 (en haut) et le Tableau 5.3.a respectivement.

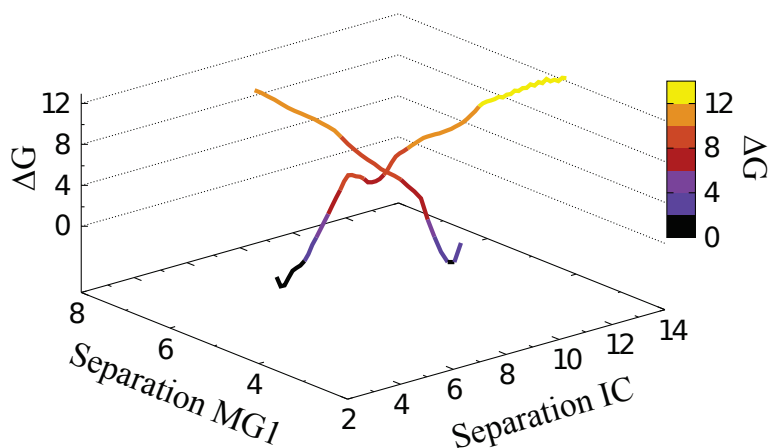
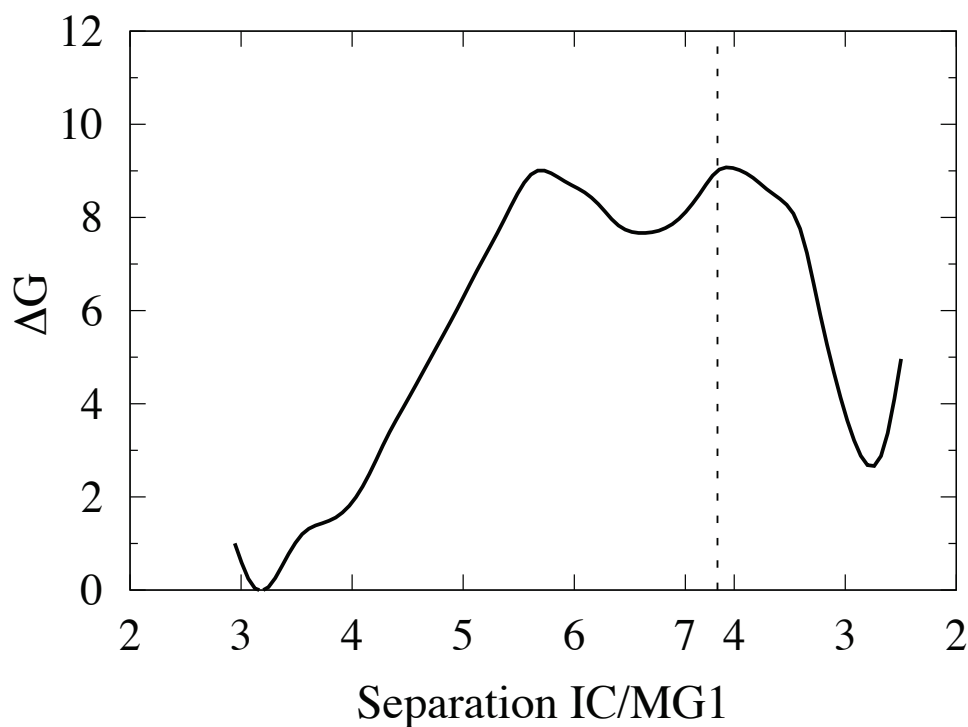


Figure 5.11 – En haut : profil d'énergie du chemin menant la DNM de son site d'intercalation à un état lié au petit sillon construit à partir de la jonction des chemins IC et MG1 (la jonction entre les deux profils étant représentée en pointillés). Une vue 3D de l'intersection des deux chemins est représentée en bas. Le gradient de couleur représentant l'énergie libre du système.

5.4 Caractérisation du passage de la DNM de son site d'intercalation au petit sillon de l'ADN par métadynamique

Les travaux de notre collaborateur Arnab Mukherjee ont permis de générer le chemin du passage de la DNM de son état intercalé vers l'état lié au petit sillon par métadynamique.

Cette méthode va permettre un échantillonnage non-équilibré d'un large espace conformationnel en ajoutant de façon régulière des potentiels répulsifs locaux à l'hypersurface énergétique. La métadynamique nécessite l'utilisation de variables collectives, qui, dans cette étude, seront au nombre de trois :

- X , qui décrit la distance entre le ligand et l'ADN ;
- Y , qui décrit le mouvement de la DNM dans la direction de l'axe de l'hélice ;
- θ , qui décrit l'orientation de la partie plane anthraquinone

X et θ correspondent aux mêmes variables que celles utilisées lors des travaux précédents de Mukherjee *et al.* [227].

Le ligand est séparé de l'ADN grâce à l'ajout de potentiels gaussiens répulsifs dans les régions de variables collectives déjà échantillonnées, forçant ainsi le système à explorer des régions de plus haute énergie. Après convergence, les potentiels Gaussien vont permettre de reconstruire la surface d'énergie libre.

Ces simulations de métadynamique sont présentées en détails dans [307] en Annexe B.

5.4.1 Surface d'énergie libre

La trajectoire de métadynamique effectuée sur 110 ns en conditions minimales de sel a permis de construire une surface d'énergie libre 2D moyennée le long de la coordonnée Y .

La Figure 5.12 qui présente ce résultat, permet de dégager cinq minima d'énergie : IC qui correspond à l'état intercalé ; IM qui correspond à un état intermédiaire ; MG1, MG12 et MG13 correspondant à trois états différents de liaisons au

petit sillon, MG1 présentant le minimum local le plus profond de ces trois état. La ligne en pointillés représente le chemin d'énergie libre le plus bas passant par ces minima d'énergie ainsi que par les états de transition associés : TS1 à TS4. Parmi les états liés au petit sillon, nous nous intéresserons uniquement à l'état MG1 dans ce propos.

La surface d'énergie libre présentée en Figure 5.12 permet de mettre en évidence différents points. Notamment, le minimum MG1 présente une différence d'énergie libre de 0,4 kcal mol⁻¹ par rapport à l'état IC. Ces résultats mettent également en lumière la présence d'un état intermédiaire IM bien défini, représenté par un minimum d'énergie de 7,3 kcal mol⁻¹. IM est atteint lorsque la DNM s'est éloignée d'environ 4,5 Å de la position IC le long de la coordonnée X et a tourné d'environ 40° le long de la coordonnée θ . Ce minimum local est entouré de deux états de transition : TS1 séparant IM de l'état IC par une barrière de 2,6 kcal mol⁻¹ par rapport à IM ; TS2 qui est supérieur de 2,4 kcal mol⁻¹ à IM, conduisant à l'état MG1.

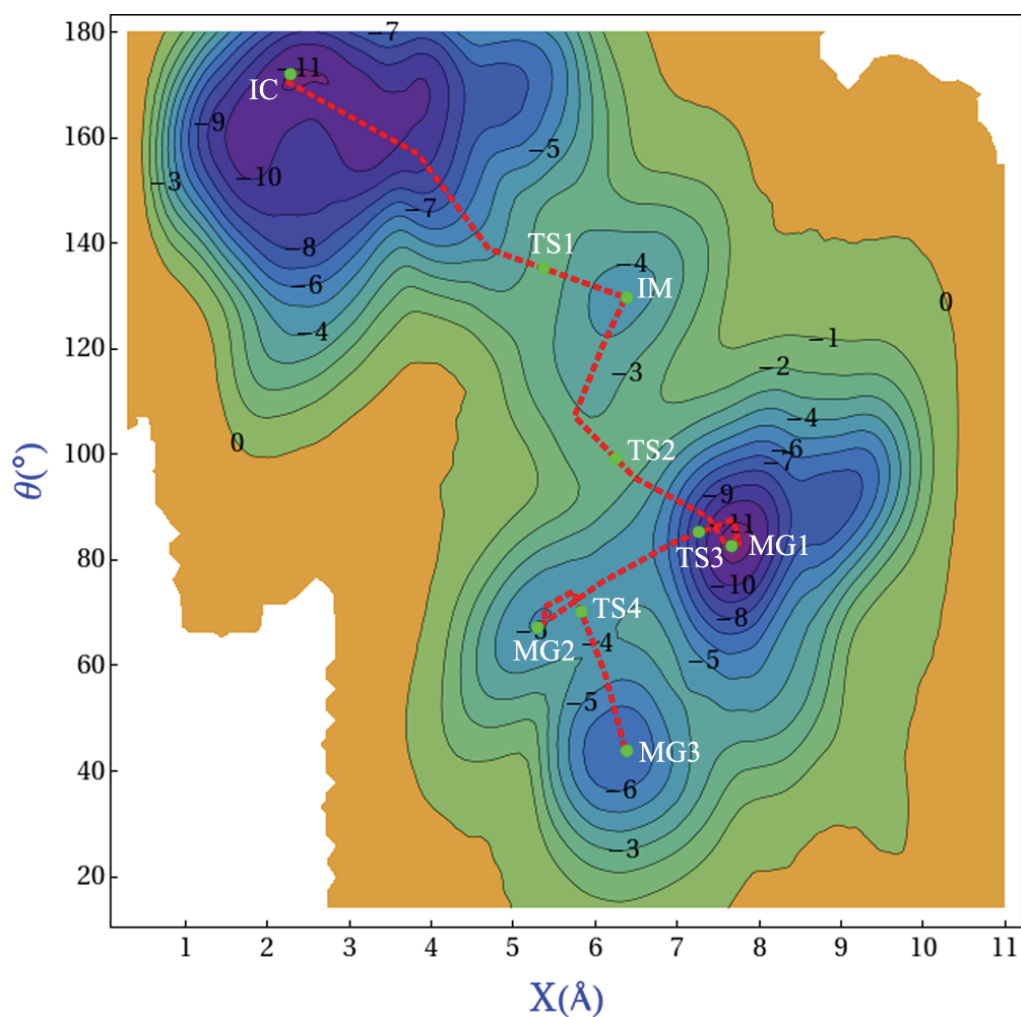


Figure 5.12 – Représentation des isosurfaces d'énergie libre échantillonnées par métadynamique en utilisant les trois variables collectives (X , θ et Y). Les valeurs le long de l'axe Y représentant le déplacement de la DNM le long de l'axe de l'ADN ont été moyennées dans le plan (X , θ). Les isosurfaces sont représentées par des lignes de niveau de 1 kcal mol⁻¹ allant du jaune au violet dans le sens de la stabilité du système. La courbe en pointillés rouge représente le chemin de plus basse énergie passant par les minima locaux les plus importants ainsi que par les états de transitions correspondants.

5.4.2 Chemin de séparation

Le chemin réalisé par métadynamique, dont les structures représentatives sont présentées par la Figure 5.14 (en bas), permet de décrire le processus de désintercalation. Le ligand entame sa sortie de la cavité d'intercalation par un mouvement le long de l'axe X tant que sa partie anthraquinone se trouve entre les paires de bases de l'ADN. Une fois sa partie plane sortie, la DNM entame une rotation (diminution de θ), qui s'accompagne d'une importante courbure de l'ADN reflétée par un roll positif important au niveau du pas C6pG7 afin de recréer l'empilement des paires de bases. Ces conformations correspondent à l'état intermédiaire IM. Par la suite, la DNM continue de lentement s'éloigner tout en conservant son mouvement de rotation, permettant à l'ADN de se relaxer et de retrouver un empilement « normal » au niveau du site d'intercalation. Finalement la DNM atteint l'état MG1 pour des valeurs de X de $\sim 7,8$ Å et de θ de $\sim 80^\circ$. Un résumé des paramètres importants des principaux états le long de ce chemin de transition est présenté dans le Tableau 5.3.b.

5.5 Comparaison des deux chemins

La comparaison des chemins de séparation créés par *umbrella sampling* et métadynamique permet de mettre en avant le fait que ces deux chemins présentent très peu de différences.

Chemin	IC	TS1	IM	TS2	MG1
Umbrella sampling	0	9,0	7,7	9,0	2,7
Métadynamique	0	9,9	7,3	9,7	0,4

Tableau 5.2 – Différence d'énergie libre (kcal mol⁻¹) des états importants par rapport à l'état intercalé le long des chemins de métadynamique et d'*umbrella sampling* dans le sens de l'état lié au petit sillon.

D'un point de vue énergétique, exception faite de l'état MG1, les énergies libres des autres états importants le long des deux chemins (résumé par le Tableau 5.2), sont très similaires : $< 1 \text{ kcal mol}^{-1}$ pour les états de transition ; $< 0,5 \text{ kcal mol}^{-1}$ pour l'état IM. Ces différences sont inférieures aux précisions des méthodes employées. L'état MG1 est $2,7 \text{ kcal mol}^{-1}$ moins stable que l'état IC le long du chemin calculé par échantillonnage parapluie, tandis que pour le chemin de métadynamique la différence d'énergie entre ces deux états n'est que de $0,4 \text{ kcal mol}^{-1}$. Cette différence d'énergie libre entre les deux chemins pour la position MG1, pourra être attribuée aux conditions ioniques qui diffèrent selon la méthode de simulation utilisée. En effet, d'après les résultats présentés dans le chapitre précédent, une différence de concentrations saline aura une influence sur les interactions interpartenaires (notamment entre l'ammonium de la DNM et le squelette de l'ADN lorsque le ligand est lié au petit sillon) et de ce fait sur l'énergie libre de dissociation du complexe.

Si l'on considère le Tableau 5.3, les deux chemins sont également proches d'un point de vue conformationnel. Les structures IC sont très similaires. De même pour les structures correspondant à l'état MG1 (mis à part au niveau du petit sillon qui sera plus large de 2 \AA pour la structure issue des simulations de métadynamique). De plus grandes différences seront cependant observables au niveau des paramètres de l'état IM ainsi que pour les deux états de transitions TS1 et TS2. Ces différences ne sont cependant pas flagrantes si l'on considère les structures de TS1 et TS2 représentées en Figure 5.14. Dans le cas de l'état IM, cette même figure met en

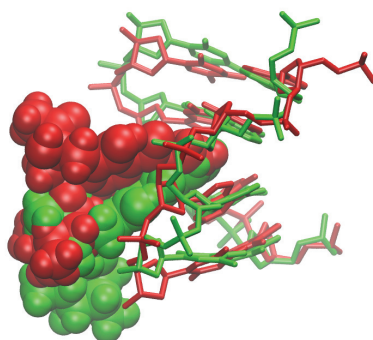


Figure 5.13 – Superposition optimale de structures de l'état IM issues des chemins d'*umbrella sampling* (en vert) et de métadynamique (en rouge).

avant une différence d'orientation du ligand. Cependant, malgré des différences au niveau des structures représentatives, les deux échantillonnages issus des deux méthodes visitent des structures très proches (illustré par la Figure 5.13). De plus, il faut également noter que les différences de structure entre les états IM concerne essentiellement le ligand ; les structures de l'ADN quant à elles, sont très similaires avec des RMSD moyenne de 1,4 Å (avec un minimum de 0,3 Å) pour les nucléotides du site d'intercalation. Finalement, la différence d'orientation de la DNM pourra en partie tirer son origine du changement de conformation interne de cycle A, ainsi que des différences de concentrations ioniques entre les systèmes. En effet, comme nous l'avons vu précédemment une différence de concentrations salines dans le milieu pourra influencer la stabilité des conformations du cycle A, et ainsi, influencer sur la position de tout le ligand.

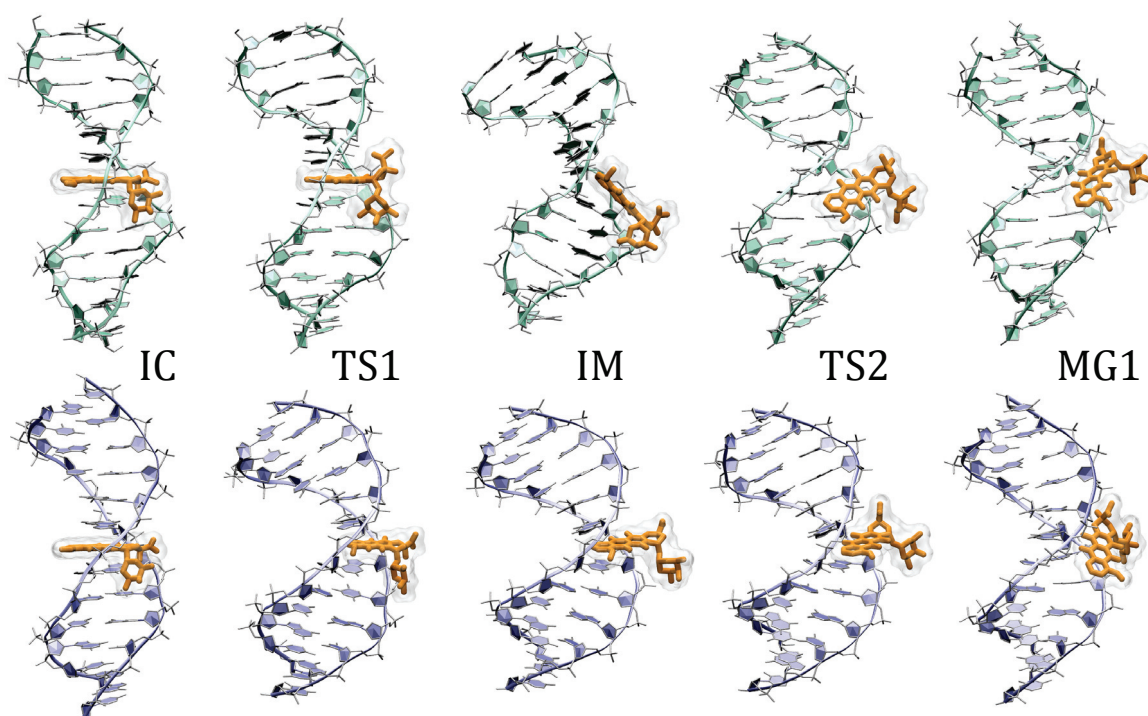


Figure 5.14 – Structures représentatives du passage de la DNM de son site d'intercalation à son site de liaison au petit sillon (de droite à gauche : IC → TS1 → IM → TS2 → MG1) pour les chemins d'*umbrella sampling* (en haut avec l'ADN représenté en vert) et de métadynamique (en bas avec l'ADN représenté en bleu).

(a) *Umbrella sampling*

Paramètres	IC	TS1	IM	TS2	MG1
Rise C6pG7 (Å)	7,4	7,2	4,9	3,6	3,8
Roll C6pG7 (°)	0	6	39	2	-4
Σ Roll A5-T8 (°)	-9	8	32	6	-4
Twist C6pG7 (°)	34	24	14	35	39
Σ Twist A5-T8	87	77	74	95	99
Largeur du petit sillon (Å)	7,0	9,0	11,3	6	4,9
Cycle A	Eq	Eq	Ax	Ax	Eq

(b) Métadynamique

Paramètres	IC	TS1	IM	TS2	MG1
Rise C6pG7 (Å)	7,4	7,3	6,0	5,5	3,8
Roll C6pG7 (°)	-2	28	44	42	8
Σ Roll A5-T8 (°)	-2	40	45	44	7
Twist C6pG7 (°)	38	25	28	29	39
Σ Twist A5-T8	81	74	84	84	99
Largeur du petit sillon (Å)	9,5	10,4	10,9	10,5	6,9
Cycle A	Eq	Eq	Eq	Ax	Ax

Tableau 5.3 – Comparaison des structures des états importants le long des chemins d'énergie libre menant la DNM de son site d'intercalation (IC) à son état lié au petit sillon (MG1) en passant par un état intermédiaire (IM). Tous les paramètres de l'ADN ont été calculés avec Curves+ [314].

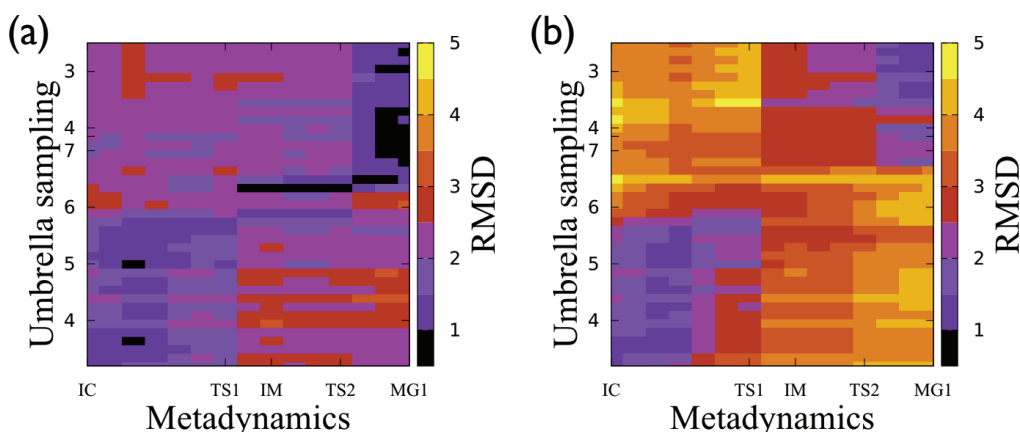


Figure 5.15 – RMSD-2D (Å) entre les conformations représentatives du chemin de métadynamique (en abscisse), et celles issues du chemin *umbrella sampling* (en ordonnées) : **(a)** pour l'ADN seul ; **(b)** pour le complexe ADN-DNM.

Afin de comparer plus globalement les deux chemins nous avons tout d'abord soumis les structures représentatives issues des deux chemins à une RMSD-2D où les structures sont comparées en fonction des atomes du complexe en entier dans un cas, et uniquement en fonction des atomes de l'ADN dans l'autre cas. Les résultats illustrés par la Figure 5.15 nous montrent une grande similarité des structures d'importance le long des deux chemins, que l'on considère seulement l'ADN (avec des $\text{RMSD} < 2 \text{ Å}$) ou bien tout le complexe ($\text{RMSD} < 2,5 \text{ Å}$). Seule la structure IM, dont le ligand est orienté différemment sur les deux chemins, montre une divergence importante des structures lorsque l'on considère le complexe ADN-DNM. La plupart des différences existant entre les deux chemins peut être attribuée aux choix avec lesquels ont été définis les structures représentatives, conséquence du fait que la détermination de ces points significatifs est limitée par la précision des courbes d'énergie libre.

Finalement, la Figure 5.16 où sont couplées les informations provenant de la structure de l'ADN au niveau du pas C6pG7 (*rise* et *roll*), de la position du ligand (distance par rapport au centre de l'axe de l'ADN), ainsi que de l'énergie libre du système (gradient de couleur), nous permet d'ajouter une vision plus continue des

chemins. Cette représentation 3D de plus de 15 000 structures issues du chemin simulé par *umbrella sampling*, nous permet de clairement localiser trois nuages de points bien distincts. Le premier, situé en bas de la figure, correspondant à l'état intercalé IC, le second (au milieu à droite) représentant les structures de l'état intermédiaire IM, et le troisième, situé en haut, l'état MG1. Les lignes noires reliant les points significatifs construits par *umbrella sampling* (cercles pleins) et métadynamique (cercles vides) nous montrent que les deux chemins sont très proches. Cette représentation ne prenant en compte que la distance entre le ligand et l'ADN, la différence d'orientation de la DNM entre les états IM ne sera donc pas visualisée ici. Cette figure nous permet également d'observer qu'en termes des variables considérées ici, les états de transition TS1 et TS2 sont placés proches de l'état IM sur le chemin de métadynamique comparé au chemin *d'umbrella sampling*. Cette différence est une conséquence directe du fait que dans ces zones de haute énergie, les transitions conformationnelles sont très rapides comme en témoigne le peu de structures reliant les différents nuages. Afin de mieux caractériser ces zones de transition, des travaux basés sur des simulations sans contraintes sont encore en cours de réalisation actuellement.

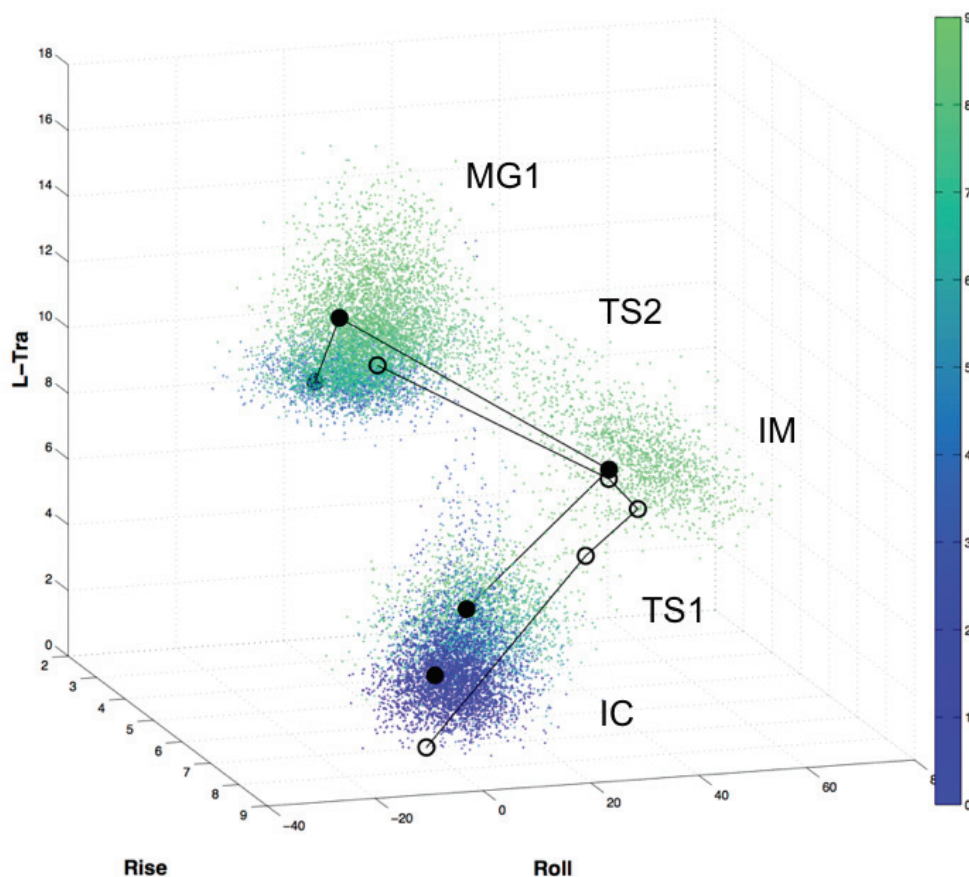


Figure 5.16 – Représentation 3D des structures issues du chemin d'*umbrella sampling* en fonction de leurs paramètres de *rise*, *roll* (dans le plan horizontal) et de la distance entre le ligand et l'axe de l'ADN (L-Tra, sur l'axe vertical). La couleur de chaque point correspond à la différence d'énergie libre de la structure qu'il représente (en kcal mol⁻¹) par rapport à l'état intercalé, allant du bleu au vert dans le sens de l'augmentation de l'énergie. Le nuage en bas du graphique correspond à l'état IC, le nuage à droite à l'état IM, et le nuage en haut correspondant à l'état MG1. Les états de transitions se trouvent dans les nuages les moins denses reliant IC et IM pour TS1, et IM à MG1 pour TS2. Les cercles pleins et vides reliés correspondent respectivement aux structures représentatives issues des chemins d'*umbrella sampling* et de métadynamique.

5.6 Discussion et conclusion

A l'aide d'une contrainte biaisant au minimum le système, nous avons pu séparer la DNM de l'ADN à partir de deux structures où le ligand est (i) intercalé entre deux paires de base, (ii) lié au petit sillon de l'ADN. La combinaison des deux chemins de dissociation créés par umbrella sampling, nous a permis de dégager une vue d'ensemble du processus d'intercalation d'un ligand entre deux paires de bases de l'ADN. Appuyé par des travaux complémentaires de métadynamique réalisés en collaboration avec Arnab Mukherjee (Université de Pune, Inde), notre modèle pour le chemin d'intercalation implique une étape initiale de liaison au petit sillon de l'ADN (MG1) (caractérisée par de faibles déformations de la double hélice ainsi que par une énergie libre de liaison de $\sim 10 \text{ kcal mol}^{-1}$), état déjà proposée précédemment dans [227]. Par la suite, ce processus va nécessiter un apport d'énergie de $\sim 6,5 \text{ kcal mol}^{-1}$ accompagné d'une rotation du ligand, lui permettant d'atteindre un état intermédiaire IM (localisé au niveau d'un puits d'énergie de 1,5 à 2 kcal mol^{-1}). A partir de ce point, seul un faible apport d'énergie sera nécessaire pour passer la barrière conduisant à l'intercalation du ligand.

D'un point de vue cinétique, ce résultat est partiellement en accord avec le modèle en trois étapes proposé par Chaires [219]. En effet, si nous assimilons notre état MG1 à l'état de liaison « extérieure » du modèle (étape 1), et notre état IC avec l'étape finale proposée par Chaires (étape 3), alors les estimations des énergies d'activation nécessaires à la DNM pour s'échapper du site d'intercalation ($14,3 \text{ kcal mol}^{-1}$), ainsi que du petit sillon ($10,7 \text{ kcal mol}^{-1}$) sont proches de nos résultats (~ 9 et $\sim 10 \text{ kcal mol}^{-1}$ pour s'échapper de IC et MG1 respectivement). Lors de nos simulations, la seconde barrière conduisant de l'état IM, qui n'apparaît pas dans le modèle de Chaires, pourrait quant à elle, contribuer à l'énergie libre associé à l'étape 1 proposée par ce modèle. Cependant, au cours de nos simulations, il ne nous a pas été permis d'observer de processus assimilables à l'étape « unimoléculaire » proposé par Chaires. Selon l'hypothèse proposé par [227], cette étape pourrait correspondre à une intercalation hors-équilibre au niveau d'un autre site au sein de l'ADN.

Au delà des aspects cinétiques, il nous semble important de nous pencher plus longuement sur l'état intermédiaire IM que nous avons pu observer. Cet état, qui n'a jamais été décrit au cours d'études précédentes concernant la DNM, mérite toute notre attention, notamment vis à vis de sa conformation. En effet, à ce niveau, le ligand se trouve partiellement intercalé entre deux paires de bases d'un ADN présentant, quant à lui une structure ouverte vers le petit sillon (et donc courbé vers le grand sillon). Des états similaires ont pu être observés dans le cadre d'interactions entre l'ADN et des molécules d'interactions au petit sillon (telles SRY [321], HMGB1 [322], Sac7d [323] ou encore des protéines se liant aux boîtes TATA [324]). Dans chacun de ces cas, la protéine intercale partiellement une chaîne latérale hydrophobe entre deux paires de bases successives, engendrant des structures d'ADN caractérisées par un *roll* positif ainsi qu'une courbure importante vers le grand sillon. Il est reconnu que de telles structures jouent un rôle dans la sélectivité et, dans certains cas, la stabilité des protéines interagissant avec le petit sillon. Ce type de déformation (produit, par exemple par le cis-Pt) peut aussi influencer la liaison d'autres protéines se liant du côté du grand sillon et conduire au blocage du processus de réparation de l'ADN puis, à l'apoptose cellulaire [322,325]. De ce fait, une stabilisation de l'état intermédiaire IM (via par exemple des modifications chimiques de la DNM ou d'autres ligands du même type), pourrait fournir une nouvelle direction intéressante de recherche thérapeutique.

Finalement ces travaux, en plus de rajouter une pierre à l'édifice de la compréhension des processus d'intercalation, permettent d'orienter ces recherches dans de nouvelles directions. Il pourrait cependant être plus qu'intéressant de compléter l'étude, en caractérisant notamment le rôle de l'eau et des ions dans la formation du complexe d'intercalation, ou encore en étudiant les mouvements de la DNM lorsqu'elle se trouve liée au petit sillon de l'ADN. Ce dernier point étant l'objet de notre dernier chapitre.

CHAPITRE 6

Glissement de la DNM le long du petit sillon de l'ADN

Une fois le noyau de la cellule atteint, avant de pouvoir s'intercaler, la DNM, au même titre que toutes les molécules interagissant avec l'ADN doit non seulement localiser la double hélice, mais également atteindre son site d'intercalation.

Les travaux que nous avons présentés dans le chapitre précédent nous permettent d'avancer de nouveaux éléments quant au mécanisme d'intercalation de la DNM entre deux paires de bases de l'ADN, sans pour autant nous apporter d'éléments concernant le processus guidant la molécule intercalante vers sa cible.

D'une manière générale, la compréhension des mécanismes guidant une molécule vers son site d'interaction sur l'ADN est encore limitée. Différentes théories ont été proposées concernant notamment la diffusion des molécules vers leur site d'action, et certains points restent controversés.

Ce dernier chapitre va s'articuler en deux parties : dans un premier temps un point sera fait sur les processus de localisation proposés actuellement, puis, dans un second temps, nous allons présenter nos travaux concernant le glissement de la DNM le long du petit sillon de l'ADN.

6.1 Un point sur les stratégies de localisation et le glissement

La localisation de son site d'action par une molécule interagissant avec la double hélice d'ADN (tels des médicaments anti-cancer, des facteurs de transcription, de régulation, etc.) ne se résume pas par une simple stratégie de diffusion 3D/collision dans le milieu. Ce phénomène a pu être mis en évidence grâce à de nombreux travaux impliquant des facteurs de transcription et notamment le répresseur Lac I [326].

Dans le cas d'une diffusion 3D la constante d'association devrait être de l'ordre de $10^8 \text{ M}^{-1} \text{ s}^{-1}$, soit la limite de diffusion [327]. Cependant de nombreux travaux expérimentaux ont montré des constantes de association supérieures à cette limite

pouvant aller jusqu'à deux ordres de grandeur (soit $10^{10} \text{ M}^{-1} \text{ s}^{-1}$) dans certains cas [326,328-331].

Ces observations ont permis d'émettre l'hypothèse que le processus 3D de recherche de cible pouvait être accéléré via un couplage à d'autres types d'évènements d'ADN [332-334].

Actuellement, il est communément admis que l'association d'un ligand avec sa cible puisse être accélérée par un phénomène de « diffusion facilitée », alternance de phases de diffusion 3D dans le milieu et de phases de diffusion quasi-1D comprenant des processus tels le glissement le long de l'ADN, de petits sauts locaux ou encore des sauts inter-segments (Figure 6.1), avec néanmoins des zones d'ombres sur le véritable impact de chacun de ces événements sur l'accélération du processus de recherche [335].

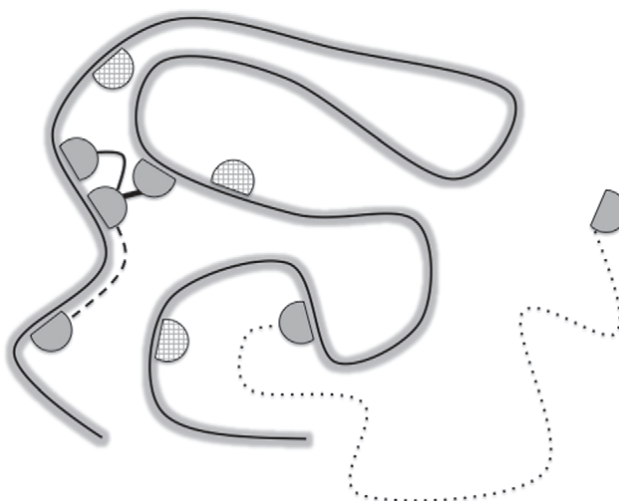


Figure 6.1 – Représentation schématique des différents mécanismes proposés pouvant être impliqués dans le processus de recherche du site d'interaction par une molécule. Les arcs circulaires gris foncé représentent une molécule cherchant sa cible sur l'ADN (représentée par la ligne noire avec l'ombre grise) par glissement (tirets noirs), petits sauts locaux (ligne noire en arc reliant deux molécules), sauts inter-segments (ligne noire en gras) ou diffusion 3D (ligne en pointillés). Les arcs circulaires grisés représentent d'autres molécules déjà attachées sur leur cible. (Extrait de Zakrzewska et Lavery [335]).

La notion de diffusion facilitée a cependant récemment été remise en question suite notamment à des travaux expérimentaux [336] ou encore de dynamique brownienne basée sur des modèles gros grains [337,338], mais sans pour autant remettre en cause le phénomène de glissement hélicoïdal le long de la double hélice, même si Halford limiterait ce glissement à des distances de l'ordre de 60-100 paires de bases maximum [336].

Dans cette partie, nous nous intéresserons uniquement au phénomène de glissement, les études consacrées aux phénomènes de sauts n'étant ici pas le sujet de notre propos.

De nombreuses études se sont focalisées sur le réel impact du glissement d'un ligand le long de l'ADN dans le processus de recherche de sa cible et ont permis de proposer différents modèles [336,339-341]. Le glissement d'un ligand le long de l'ADN pourra être soit linéaire avec un mouvement de diffusion translationnelle, soit suivre les sillons de l'ADN et de ce fait, adopter une trajectoire hélicoïdale permettant de maintenir l'interface ligand-ADN tout au long du mouvement. Dans le cas de la seconde hypothèse, la diffusion est cependant ralentie comparée au premier cas, et fortement dépendante de la taille du ligand (soit pour un ligand de rayon r , de l'ordre de r^{-3} dans le cas d'un mouvement hélicoïdal au lieu de r^{-1} dans le cas d'une diffusion translationnelle). Les travaux expérimentaux menés par Blainey en 2009 [340] ont permis de valider la seconde hypothèse.

Appuyé par les modèles simples, le glissement sera plus rapide que la diffusion 3D à condition que la profondeur des puits rencontrés sur la surface d'énergie libre n'excède pas 1 à 2 k_bT (soit ~ 1 kcal mol $^{-1}$). Ces modèles suggèrent également qu'une recherche optimale nécessite qu'un ligand passe un temps équivalent en phase de recherche 1D qu'en phase de recherche 3D [339,342-345].

A l'heure actuelle, seules quelques rares études ont visé à recréer le glissement d'un ligand le long de l'ADN par des techniques de simulation moléculaire. On pourra notamment citer les travaux de Givaty et *al.* [346], basés sur des simulations de dynamique brownienne utilisant un modèle gros grain, ou encore ceux de Chen et *al.* [347], combinant dynamique brownienne et dynamique moléculaire.

Cette dernière étude a notamment permis d'obtenir des barrières de diffusion de l'ordre de 1 kcal mol^{-1} mais avec de nombreuses approximations concernant les énergies d'interactions utilisées.

Récemment, une étude de dynamique moléculaire visant à faire glisser un domaine de liaison du répresseur Lac sur un brin d'ADN en utilisant une représentation tous atomes des systèmes a permis d'obtenir des barrières de l'ordre de 9 kcal mol^{-1} [348]. Cependant, ces travaux ont nécessité d'une part le gel de la conformation de l'ADN, mais également l'emploi d'une structure de la double hélice parfaitement symétrique, pour pouvoir définir un chemin de glissement suivant rigoureusement une hélice régulière.

On pourra finalement citer les travaux de DM menés par Vargiu et *al.* [349], qui ont permis de faire glisser deux agents alkylants de l'ADN le long du petit sillon de l'ADN avec des barrières d'énergie de l'ordre de 6 kcal mol^{-1} . Cependant, cette étude est limitée par la méthode employée qui consiste à contrôler simplement la distance entre un point ancré dans l'ADN et un point au sein du ligand. Une telle coordonnée réactionnelle ne permettra pas de garantir un chemin passant par le petit sillon de l'ADN, en plus de potentiellement induire d'autres déformations que le simple glissement du ligand.

6.2 Stratégie et contrainte de glissement

Afin d'augmenter notre compréhension du processus guidant la DNM jusqu'à son site d'intercalation, il nous est donc paru évident de tenter de simuler le glissement de cette dernière le long du petit sillon de l'ADN. Pour cela, il nous a fallu envisager une contrainte de glissement répondant à un certain nombre de critères tels (i) la conservation de la flexibilité de la double hélice ; (ii) permettre un mouvement hélicoïdal tout en laissant la possibilité à la DNM de pouvoir se réorienter librement dans le sillon ; (iii) éviter au maximum les déformations de la double hélice engendrées par la contrainte elle-même, tout en biaisant le système au minimum.

À partir de ces critères, nous avons développé une contrainte permettant de définir la position du ligand dans le petit sillon indépendamment des déformations pouvant avoir lieu au sein de la double hélice :

Soit le point \mathbf{p}_i , milieu du segment reliant deux atomes équivalents (préalablement choisis) sur les brins Watson et Crick au niveau de la paire de base i , et respectivement dénotés a_{iw} et a_{ic} :

$$\mathbf{p}_i = \frac{a_{iw} + a_{ic}}{2} \quad (6.72)$$

Dans cette étude nous avons choisi les atomes C1' des sucres désoxyribose pour définir les atomes a_i . On en déduit les vecteurs unitaires \mathbf{q}_i selon :

$$\mathbf{q}_i = \frac{(\mathbf{p}_i - \mathbf{p}_{i-1})}{|\mathbf{p}_i - \mathbf{p}_{i-1}|} \quad (6.73)$$

A partir de cette construction, il est possible de combiner les vecteurs \mathbf{q}_i pour obtenir un chemin curviligne constitué de vecteurs u_d (passant par des points v_d) définissant le centre du petit sillon. Par la suite, pour $i-1 < d \leq i$ un point intermédiaire v_d est défini tel que :

$$v_d = (i - d) \mathbf{p}_{i+1} + (d - i + 1) \mathbf{p}_i \quad (6.74)$$

puis,

$$\begin{cases} \mathbf{u}_d = (i - d - 0.5) \mathbf{q}_{i-1} + (d - i + 1.5) \mathbf{q}_i & d < i - 0.5 \\ \mathbf{u}_d = (i - d + 0.5) \mathbf{q}_i + (d - i + 0.5) \mathbf{q}_{i+1} & d \geq i - 0.5 \end{cases} \quad (6.75)$$

Finalement, pour contraindre un ligand, représenté par un point r (correspondant à un atome spécifique ou au centre de masse d'un groupe d'atome) à une position d le long du petit sillon, il suffit de calculer le point v_d et le vecteur u_d et d'imposer une contrainte quadratique sur la valeur s correspondant à l'écart de la posi-

tion du ligand par rapport à v_d , en termes de sa projection sur le vecteur u_d (voir Figure 6.2 à droite), soit :

$$s = (r - v_d) \cdot \left(\frac{u_d}{|u_d|} \right) \cdot \frac{1}{|p_i - p_{i-1}|} \quad (6.76)$$

Par la suite, le potentiel appliqué sur le système sera de la forme :

$$E = ks^2 \quad (6.77)$$

où k correspond à la constante de force prédéfinie.

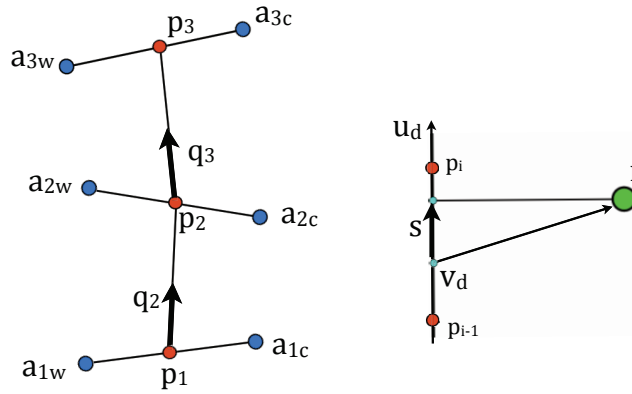


Figure 6.2 – Représentation schématique des différentes composantes de la contrainte utilisée pour faire glisser la DNM le long du sillon de l'ADN.

Ainsi, seule la position du ligand le long d'un fragment d'ADN (en unités de paires de bases) sera contrainte, sans influencer son orientation, sa distance par rapport à l'ADN, ou sa conformation interne. On note que la valeur d mesure la position du ligand en unités de paires de bases et non en Angströms.

6.3 Validation du modèle

Avant d'utiliser cette contrainte sur un système hydraté composé de nombreuses molécules d'eau, nous avons tout d'abord validé l'efficacité de notre outil sur un système simple.

Pour cela nous avons simulé le glissement de la DNM sur un oligomère d'ADN figé et de séquence régulière *in vacuo*.

Un tel système nous permet de nous affranchir du « bruit » induit par la plasticité de l'ADN en solvant aqueux, ainsi que de potentiels effets de séquences.

6.3.1 Conditions de simulations et étapes préliminaires

La contrainte a été testée sur deux complexes ADN/DNM composés d'un tétrakaidecamère d'ADN-B canonique, de séquence polyG, généré puis relaxé par minimisation d'énergie avec JUMNA, tout en conservant une symétrie hélicoïdale. Les deux systèmes se différencient par la distance inter-paire de bases : dans le premier cas elle est constante tout au long de la séquence (soit 3,40 Å ; tandis que dans le second cas les valeurs de *rise* sont alternées régulièrement entre 3,20 Å et 3,60 Å. Dans les deux cas, la DNM est ensuite placée à proximité du centre de la séquence. Afin d'écartier une influence éventuelle de l'asymétrie du ligand, le glissement a été réalisé dans le vide en partant du centre de la séquence vers l'extrémité 3' puis dans le sens opposé, soit vers l'extrémité 5' de la double hélice. Pour ces simulations, tous les atomes lourds des doubles hélices ont été fixés dans l'espace. Cette fixation, en plus d'éliminer l'influence de la flexibilité de l'ADN, permet d'éviter la désagrégation de l'ADN en l'absence de solvant.

Dans un premier temps, les systèmes ont été minimisés puis progressivement montés en température jusqu'à atteindre 300 K avant d'être simulés pendant 5 ns

afin d'équilibrer le système. Toutes ces étapes ont été réalisées avec une constante diélectrique $\epsilon = 20$.

Dans un second temps, le glissement a été caractérisé par *umbrella sampling*. Pour cela la contrainte a été appliquée sur l'atome C18 de la DNM, tandis qu'au niveau de l'ADN, les C1' ont été choisis comme références pour définir les vecteurs p_i .

La valeur de la coordonnée réactionnelle (ξ) de la première fenêtre d'échantillonnage correspond à la position de l'atome C18 (par rapport à l'ADN) de la dernière structure issue de la simulation de relaxation.

Après 2 ns d'échantillonnage initial, nous avons réalisé 140 fenêtres de 2 ns de simulations par pas de 0.05 paires de bases, en imposant une constante de force de 800 kcal mol⁻¹.Å⁻². Les structures et vitesses initiales de chaque fenêtre correspondant aux structures et vitesses du dernier pas de dynamique de la fenêtre précédente. Les PMFs issus de ces simulations ont ensuite été calculés par WHAM en retirant les premiers 25 % de simulation (500 ps) de chaque échantillonnage.

6.3.2 Séquence polyG régulière

Dans le cas du système avec une séquence polyG régulière, la contrainte de la fenêtre de départ a été fixée à une valeur de départ $\xi = 7,85$. En variant la valeur de cette contrainte, la DNM a ensuite pu être déplacée dans le sillon sur une distance de plus de 3 paires de bases vers chaque extrémité du double brin.

Le PMF résultant de ce glissement (voir Figure 6.3) nous montre une parfaite périodicité au niveau des énergies le long de la coordonnée réactionnelle, et ce indépendamment du sens de simulation. Cette périodicité est équivalente à une paire de bases. D'un point de vue structurel, dans les deux sens de simulation, la DNM reste parfaitement liée au sillon.

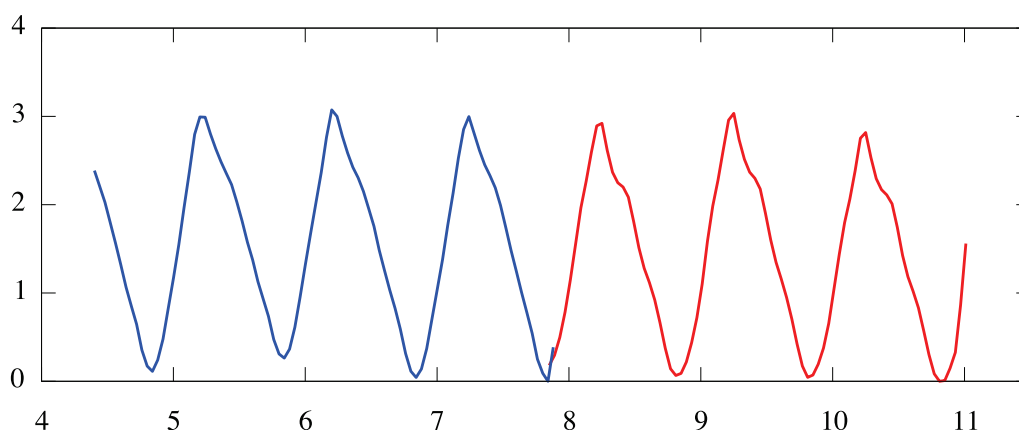


Figure 6.3 – Profil d'énergie libre le long du chemin de glissement de la DNM dans le sillon d'un ADN de séquence polyG canonique fixée. En bleu, glissement en direction de l'extrémité 5', en rouge, glissement vers l'extrémité 3'.

6.3.3 Séquence polyG avec *rise* alterné

Pour le système dont les valeurs de *rise* des bases ont été alternées, la position de la DNM est contrainte à proximité de la 8^{ème} paire de base (7,9 exactement). Au cours de ces simulations, la DNM a pu être glissée de plus de deux paires de bases en direction de chaque extrémité de l'ADN.

Le profil d'énergie libre (Figure 6.4) issu de ces simulations est différent des résultats précédents notamment au niveau de la périodicité des variations des énergies. En effet, dans ce cas la périodicité du profil est équivalente à deux paires de bases (alors qu'elle était de 1 dans le cas précédent). Ce changement de périodicité est parfaitement corrélé avec les variations des distances inter-paires de bases le long de la séquence. Les pics de plus haute énergie localisés entre deux paires de bases distantes de 3,6 Å l'une de l'autre, tandis que les pics présentant une énergie

inférieure se trouvent au niveau des bases séparées par un rise de 3,2 Å. De même que dans le cas précédent, le sens de glissement n'a aucune influence sur l'énergie libre du système et la DNM reste parfaitement placée dans le sillon.

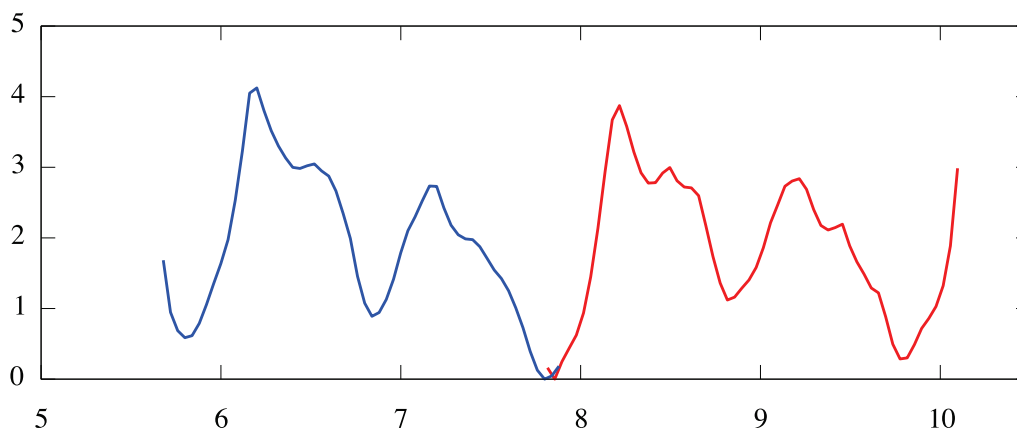


Figure 6.4 – Profil d'énergie libre le long du glissement de la DNM dans le sillon d'un ADN de séquence polyG fixée dont les valeurs de rise sont alternées. En bleu, glissement en direction de l'extrémité 5', en rouge, glissement vers l'extrémité 3'.

Les résultats issus de ces deux simulations nous permettent de valider la contrainte utilisée. En effet, dans les deux cas les résultats obtenus montrent un accord parfait entre la périodicité du profil d'énergie libre associé au glissement et la périodicité spatiale du double brin d'ADN.

Cependant il est important de rappeler que le modèle utilisé pour ce test préliminaire ne reflète en rien la réalité aussi bien d'un point de vue physique que physiologique. Par conséquent, ces simulations ne nous permettent d'avancer aucune conclusion concernant l'éventualité de l'existence d'un processus de glissement de la DNM le long du petit sillon de l'ADN, même si elles nous permettent d'envisager l'utilisation de cette contrainte dans un modèle plus réaliste.

6.4 Simulation du glissement de la DNM dans l'eau

La suite logique de nos travaux est donc la caractérisation du glissement de la DNM en se basant sur un modèle avec une représentation explicite de l'eau, une concentration physiologique en ions ainsi qu'un ADN flexible.

6.4.1 Description du système

Dans ce sens, nous avons entrepris de simuler le déplacement de la DNM le long du sillon du dodécamère de séquence $d(\text{GCGCACGTGCGC})_2$ utilisé dans nos travaux précédent.

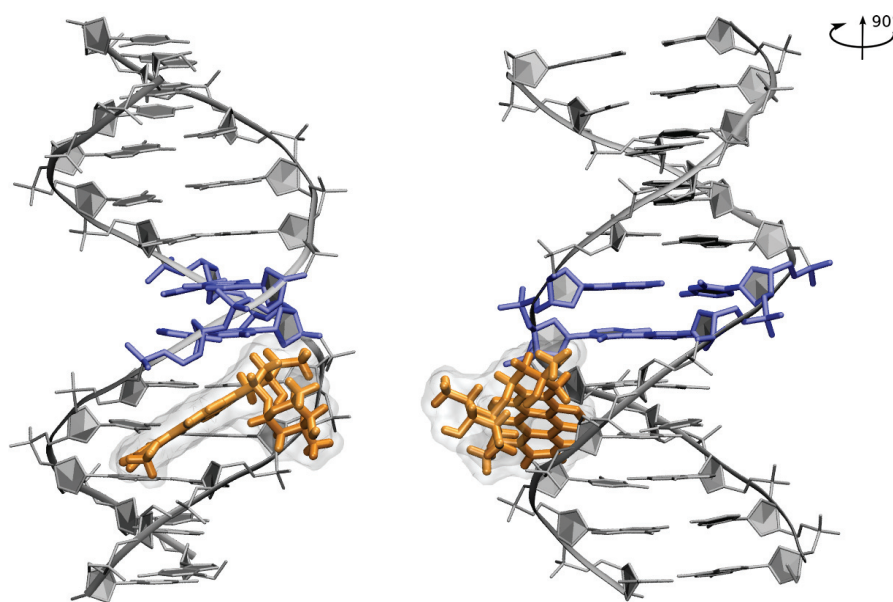


Figure 6.5 – Représentation de la structure de départ ayant servi à simuler le glissement de la DNM le long du sillon de l'ADN. La DNM (représentée en orange) est placée entre l'extrémité 5' et le site d'intercalation du ligand (en bleu).

Construction / Equilibration : A partir de l'oligomère double brin construit avec JUMNA, la partie rigide de la DNM est placée dans le sillon entre l'extrémité 5' et le centre du fragment où est localisé le site d'intercalation du ligand. Le ligand est orienté de façon à ce que son cycle flexible placé dans le sillon pointe vers l'extrémité 3' de la séquence, le sucre daunosamine étant quant à lui en dehors du sillon avec l'ammonium dirigé vers le squelette phosphaté du brin Crick (Figure 6.5).

Ce placement, hors du centre de l'oligomère, est guidé par la nécessité de caractériser le déplacement du ligand sur un grand nombre de paires de bases, ainsi que par la mise en avant d'un éventuel effet de séquence lors du passage de la DNM au niveau du site d'intercalation.

La construction obtenue a ensuite été placée dans une boîte cubique, hydratée avec de l'eau TIP3P (~7400 molécules), puis amenée à une concentration physiologique en ions Na^+ et Cl^- . Ces étapes de construction/hydratation, ainsi que celles de minimisation d'énergie et de relaxation du système ont été réalisées suivant le même schéma que dans les parties précédentes (voir partie 4.1). Le système est finalement simulé sans contraintes pendant 5 ns afin de l'équilibrer.

Les coordonnées spatiales et les vitesses de la dernière structure de l'équilibration servent de point de départ à nos simulations sous contrainte.

Umbrella sampling : Afin de déplacer le ligand, les atomes sur lesquels la contrainte est appliquée sont les mêmes que ceux définis lors du test décrit ci-dessus : (a) les paires d'atomes C1' au niveau de chaque paire de bases pour l'ADN, (b) l'atome C18 placé au centre du cœur aglycon pour la DNM. Le choix de cet atome a été guidé par la volonté de laisser un maximum de liberté de réorientation à la DNM, tout en minimisant une potentielle influence de la contrainte sur les conformations internes du ligand.

La coordonnée réactionnelle ξ correspond donc à la position de l'atome C18 de la DNM par rapport aux paires de bases de l'ADN.

A partir de la structure initiale, nous avons construit un chemin réactionnel déplaçant le ligand entre la position de départ (où $\xi = 4,7$) et la 9^{ème} paire de bases de la séquence (Figure 6.6). Ce chemin initial est constitué de 87 fenêtres de 2 ns d'échantillonnage par pas de 0.05 paire de bases. La structure de départ de la fenêtre $n+1$ correspondant à la dernière structure de la fenêtre n . Les échantillonnages de chaque fenêtre ont ensuite été prolongés jusqu'à un total de 4 ns.

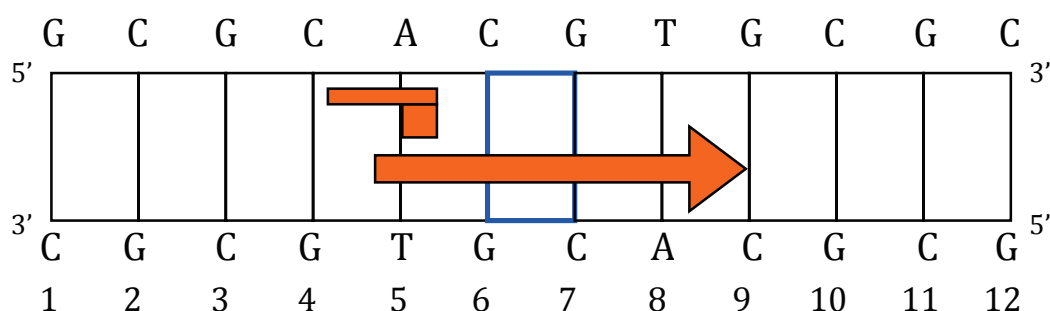


Figure 6.6 – Schéma représentatif du glissement de la DNM le long du sillon de l'ADN dans la direction $5' \rightarrow 3'$. Le déplacement de l'atome C18 de la DNM le long de la séquence est représenté par la flèche orange (allant de 4,7 à 9 paires de bases), le brin Watson étant placé en haut et le brin Crick en bas. Le site d'intercalation au niveau du pas CpG est représenté par un rectangle bleu. La DNM étant schématisée sous une forme rectangulaire (correspondant à sa partie conjuguée), associé à un carré (représentant son sucre).

6.4.2 Chemins initiaux

Dans la première version de la contrainte que nous avons utilisée pour construire initialement le chemin de glissement, l'écart entre la projection du ligand sur u_d et la position de référence v_d (soit s) était exprimé en Angströms. La valeur de

référence v_d quant à elle est exprimée en paires de bases. Dans le cas de distances entre deux points p_i et p_{i+1} régulières (soit des valeurs de *rise* constantes) tout au long du chemin de glissement, la valeur de s pourra être ramenée en unité de paires de bases par simple division par une constante. Cependant l'ADN est un objet flexible dont les distances inter-paires de bases pourront varier tout au long des simulations, et ce, que ce soit au sein d'une même fenêtre mais aussi et surtout entre deux pas de séquences différentes. L'analyse du *rise* des pas à proximité desquels le ligand était placé tout au long du chemin, a montré d'importantes variations d'une fenêtre à l'autre, variations encore amplifiées lorsque l'on changeait de pas (résultats non présentés). De telles variations, notamment au niveau des changements de pas, peuvent introduire un biais dans le calcul de s , rendant le PMF difficilement interprétable.

Afin de s'affranchir de cet effet, nous avons modifié la contrainte en ajoutant un facteur permettant de normaliser la valeur de la coordonnée réactionnelle par rapport à la distance inter-paire de bases (représenté par le 3^{ème} facteur de l'équation 6.76).

A partir des simulations initiales, nous avons ajouté 4 ns d'échantillonnages par fenêtres le long du chemin précédent avec la nouvelle version de la contrainte. Malgré la modification apportée à la contrainte, le recouvrement des échantillonnages le long de ξ reste très satisfaisant. Le profil d'énergie libre obtenu étant présenté ci-dessous.

6.4.2.1 Profil d'énergie libre

Le PMF (Figure 6.7) montre une augmentation d'énergie de plus de 16 kcal mol⁻¹ entre le point de départ du chemin et son point d'arrivée 4 paires de bases plus loin. Plus précisément, on peut notamment noter une forte augmentation d'énergie (~ 12 kcal mol⁻¹) pour $4,7 < \xi < 6,2$; suivie par un minimum local (centré sur $\xi = 6,5$), lui-même séparé par une barrière de ~ 2 kcal mol⁻¹ d'une zone où l'énergie est plus stable (entre 6,8 et 7,6). Une seconde barrière de ~ 5 kcal mol⁻¹ permet au ligand de rejoindre le pas T8pG9. Au regard du mécanisme du glissement issu de ces simulations, qui ne semble pas induire de déformations majeures et irréversibles de la double hélice, de telles énergies ne sont pas réalistes.

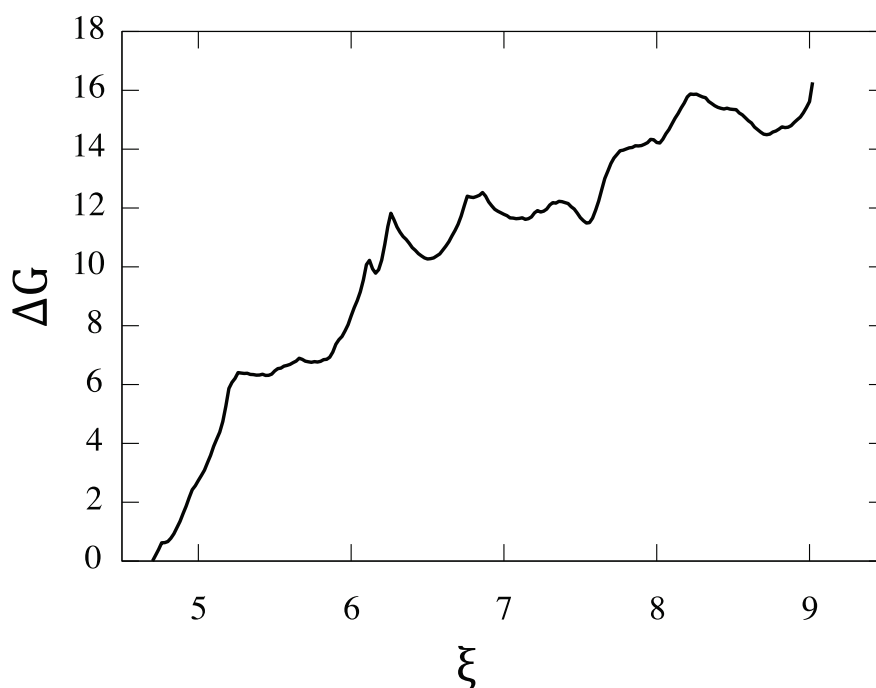


Figure 6.7 – Profil d'énergie libre du glissement de la DNM le long de 4 paires de bases.

En effet, si l'on se réfère aux études théoriques, nous nous trouvons dans une situations où la DNM est incapable de glisser puisque les barrières d'énergie pour

glisser sont bien supérieures aux $1 - 2 \text{ k}_\text{B}T$ (soit $\sim 0,6 - 1,2 \text{ kcal mol}^{-1}$) attendues pour ce type de mouvement [339,340]. Comparées aux simulations de glissement que nous avons précédemment décrites, les énergies obtenues sont quasiment deux fois plus élevées que celles observées par [348], et ~ 3 fois supérieures à celles calculées par [349].

Au vu de ces résultats, qui semblent écarter toutes possibilités à la DNM de glisser le long du petit sillon, nous nous sommes tout d'abord penchés sur la pertinence de nos échantillonnages, avant de rechercher plus en détail une quelconque raison structurale qui nous aurait échappée lors de notre visualisation (pourtant exhaustive) du mécanisme de glissement.

6.4.2.2 Cohérence des échantillonnages

Une des raisons pour lesquelles on observe d'importantes barrières d'énergie (notamment pour des valeurs de $\xi < 6,3$), pourrait être imputée à un échantillonnage insuffisant des degrés de liberté orthogonaux à la coordonnée réactionnelle choisie.

En effet, lorsque ξ est fixée à une valeur donnée, un système multidimensionnel aura également la possibilité d'échantillonner différents états métastables ou minima locaux, séparés par de hautes barrières d'énergies dans des directions orthogonales à ξ . Un échantillonnage insuffisant de telles coordonnées, peut être à l'origine de l'augmentation de l'erreur systématique dans le calcul de l'énergie libre aboutissant à une hystérèse [305]. Dans une telle situation, la trajectoire d'une seule fenêtre de simulation ne montrera aucune indication de l'échantillonnage insuffisant le long de coordonnées orthogonales. Cependant, si ces états orthogonaux sont visités lors de l'échantillonnage de fenêtres adjacentes, le problème pourra être mis en lumière en comparant le résultat des simulations concernées. Malgré cela, cette comparaison sera limitée dans le cas de systèmes biomoléculaires complexes, avec un nombre important de degrés de libertés.

Partant de ces observations, Zhu et Hummer ont développé une méthode permettant d'outrepasser cette limitation en caractérisant la cohérence des échantillonnages le long d'un chemin réactionnel [305] :

Si différents états d'une coordonnée orthogonale sont échantillonnés dans deux fenêtres adjacentes, alors les probabilités de distributions des deux fenêtres selon ξ seront incohérentes entre elles. De ce fait, la comparaison de la cohérence des histogrammes entre deux fenêtres voisines pourra permettre de mettre en lumière un éventuel échantillonnage insuffisant d'un degré de liberté orthogonal. Dans ce sens, si l'on considère les probabilités de distribution non biaisées de deux fenêtres adjacentes $p_1^{nb}(\xi)$ et $p_2^{nb}(\xi)$ le long de ξ , on pourra introduire un « coefficient d'incohérence » $\theta_{1,2}$ tel que :

$$\theta_{1,2} = \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \max \left| \int_{-\infty}^{\xi} [p_1^{nb}(\xi) - p_2^{nb}(\xi)] dx \right| \quad (6.78)$$

avec N_1 et N_2 correspondant au nombre d'échantillonnages indépendants des fenêtres 1 et 2 respectivement (voir [305] pour plus de détails).

Une valeur de $\theta_{1,2}$ anormalement haute sera le reflet d'une incohérence entre les échantillonnages des deux fenêtres testées. Le calcul de ce coefficient pourra être étendu à toutes les paires de fenêtres adjacentes.

A partir de cette méthode, nous avons donc caractérisé la cohérence des échantillonnages voisins le long du chemin de glissement.

Le calcul des valeurs de θ entre deux fenêtres d'échantillonnages adjacentes (Figure 6.8), nous montre des hautes valeurs d'incohérences pour les fenêtres d'échantillonnage centrées sur une coordonnée réactionnelle inférieure à 7. Les valeurs de ξ pour lesquels θ est le plus haut, sont corrélées avec de fortes augmentations de l'énergie libre le long du PMF illustré par la Figure 6.7. Les échantillonnages au-delà de $\xi = 7$, montrent une meilleure cohérence entre eux ($\theta < 2$), ce qui semble se traduire par des fluctuations d'énergie moins importantes.

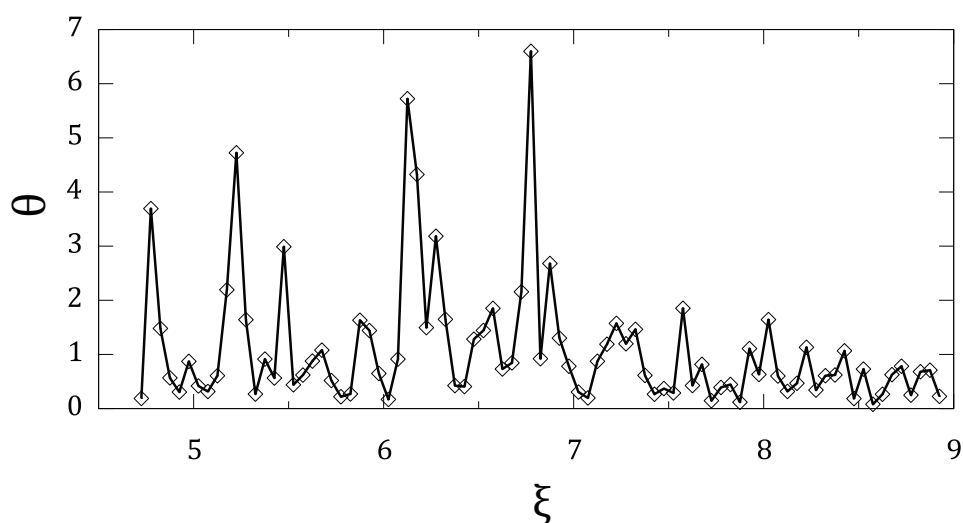


Figure 6.8 – Représentation des valeurs du coefficient d'incohérence θ le long de la coordonnée réactionnelle ξ .

Au vu de ces résultats, il nous apparaît clairement que nos échantillonnages sont insuffisants le long de ξ . Une des solutions pourrait être d'augmenter de façon importante l'échantillonnage de chaque fenêtre de simulation le long du chemin réactionnel.

Cependant, notre système présente un très grand nombre de degrés de libertés, aussi bien au niveau de l'ADN que du ligand, et il apparaît que dans le cas du glissement, ces degrés de libertés interviennent de façon non négligeable dans les échantillonnages. Par conséquent, il n'est pas à exclure qu'il soit impossible de visiter tous les états des coordonnées orthogonales à ξ dans des temps d'échantillonnages accessibles à l'heure actuelle.

Dans ce sens, afin d'améliorer la qualité des échantillonnages nous avons choisi de réduire les degrés de libertés du système en contraignant la flexibilité interne de la DNM.

6.4.3 Chemin avec la conformation interne de la DNM fixée

A partir des échantillonnages précédents, 5 ns de simulations ont été ajoutées à chaque fenêtre avec la DNM fixée dans la même conformation pour toutes les fenêtres de simulation.

Au cours de nos travaux précédents, nous avons pu caractériser la liaison de la DNM au petit sillon de l'ADN. Cette étude a permis de mettre en avant un état très stable, l'état MG1, où la DNM est placée au fond du petit sillon de l'ADN. Cette structure correspond à un minimum local ; à ce point, la DNM doit donc présenter une conformation interne stable et adaptée au petit sillon de la double hélice. En conséquence de quoi, la structure de la DNM de MG1 a servi de structure de référence lors des simulations avec une DNM de conformation fixe.

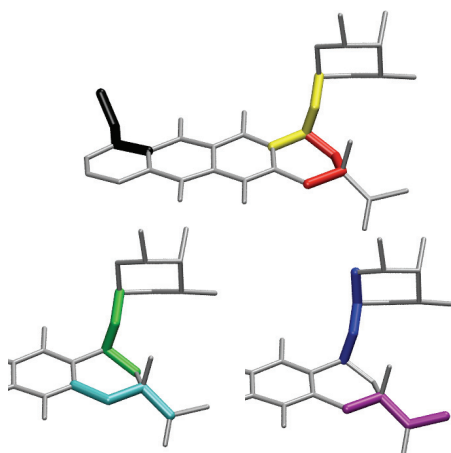


Figure 6.9 – Représentation de la structure de la DNM utilisée comme conformation de référence. Les dièdres de couleurs représentent tous les angles de torsion contraints pour fixer la conformation interne du ligand.

Pour cela, à partir de la dernière structure issue des échantillonnages décrits dans la partie 6.4.2, la structure de la DNM de chaque fenêtre est modifiée pour correspondre à la structure de référence. Plus précisément, les transitions confor-

mationnelles du ligand ont été effectuées en obligeant les principaux angles dièdres du ligand à atteindre une valeur cible (soit celle de la structure de référence dans notre cas – voir Figure 6.9) de façon progressive sur une durée de 1 ns. Cette modification graduelle de la conformation du ligand permet au système de pouvoir s'adapter au fur et à mesure des transitions. Les systèmes ont ensuite été simulés 4 ns supplémentaires avec la DNM bloquée.

Le profil d'énergie libre résultant de ces nouvelles simulations (Figure 6.10) a été calculé avec WHAM à partir des deux dernières ns d'échantillonnages de chaque fenêtre (les 2 ns précédentes étant assimilées à une phase d'équilibration des systèmes).

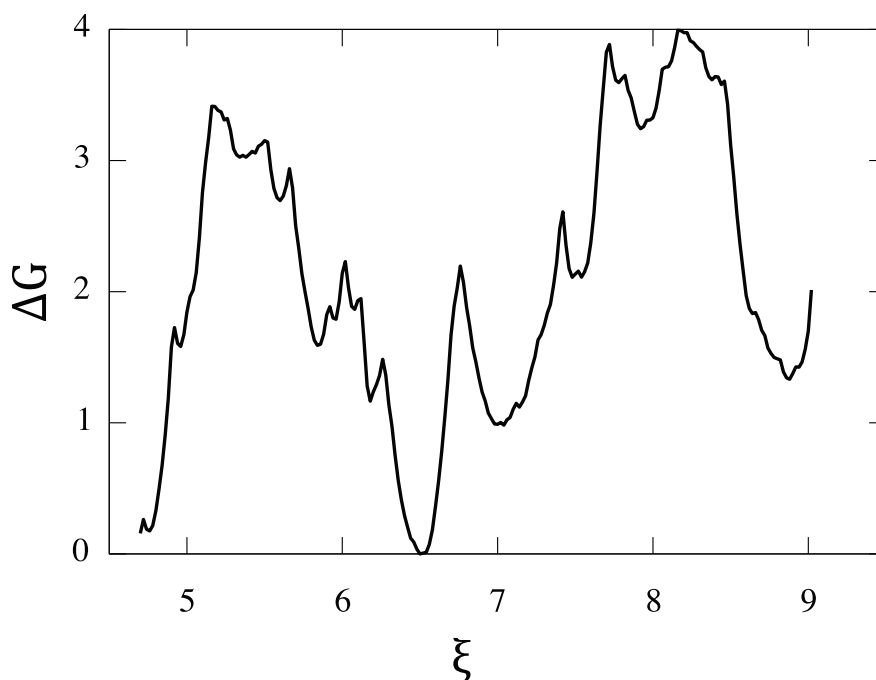


Figure 6.10 – Profil d'énergie libre (kcal mol^{-1}) du glissement de la DNM dont la conformation interne est fixée, le long du petit sillon de l'ADN.

La barrière de 12 kcal mol^{-1} observée sur le profil précédent entre la position initiale et le 6^{ème} pas a disparu pour laisser place à une barrière de $\sim 3 \text{ kcal mol}^{-1}$ dont le maximum est atteint pour $\xi = 5,2$. Ensuite, le système glisse vers le mini-

mum du profil situé au centre des 6^{ème} et 7^{ème} bases de la séquence, soit le pas C6pG7. La DNM doit ensuite passer une barrière de 2 kcal mol⁻¹ avant de rejoindre un second minimum local au niveau de la 7^{ème} paire de bases. Le ligand franchit ensuite une dernière barrière pour atteindre un plateau situé au maximum d'énergie (soit 4 kcal mol⁻¹) au niveau de la 8^{ème} paire de bases, avant de finalement glisser vers un dernier minimum centré sur la 9^{ème} paire de bases.

La réduction des degrés de liberté du système nous a permis d'améliorer l'échantillonnage conformationnel. Ce changement conduit à une réduction significative des barrières d'énergie le long du chemin de glissement, pour finalement atteindre un écart d'énergie libre maximale de 4 kcal.mol⁻¹ au lieu de 16 kcal.mol⁻¹ pour le profil précédent. Une telle diminution de l'énergie révèle clairement une meilleure équilibration des degrés de liberté non contraints restants du système.

Il est intéressant de noter l'apparition du puits d'énergie important centré sur le pas C6pC7, tandis que sur le chemin précédent, seul un petit minimum local était situé à ce niveau. En effet, ce pas correspond au site d'intercalation de la DNM. La présence de ce puits pourrait potentiellement indiquer que le ligand, en glissant le long du sillon reconnaît sa séquence cible. Cependant, il faut rappeler que pour ces simulations, la conformation du ligand est bloquée dans une conformation adaptée au petit sillon au niveau du site d'intercalation, ce qui pourrait en partie favoriser l'apparition d'un tel puits.

Lorsque l'on s'intéresse à la cohérence des échantillonnages voisins (Figure 6.11), on observe une amélioration de la cohérence des simulations adjacentes dans la partie centrale du chemin de glissement (pour $6,2 < \xi < 7,5$). Pour des valeurs de $\xi < 6,2$, l'introduction de contraintes harmoniques sur la structure de la DNM semble avoir amélioré la cohérence entre de nombreux voisins, pouvant expliquer ainsi le net changement observé dans l'énergie libre. Cette correction reste cependant limitée dans cette zone, notamment au niveau des échantillonnages centrés sur $\xi = 4,8$ et dans une moindre mesure ceux centrés sur $\xi = 5,7$ et $6,1$. Finalement, la fixation de la structure du ligand semble avoir tendance à augmenter

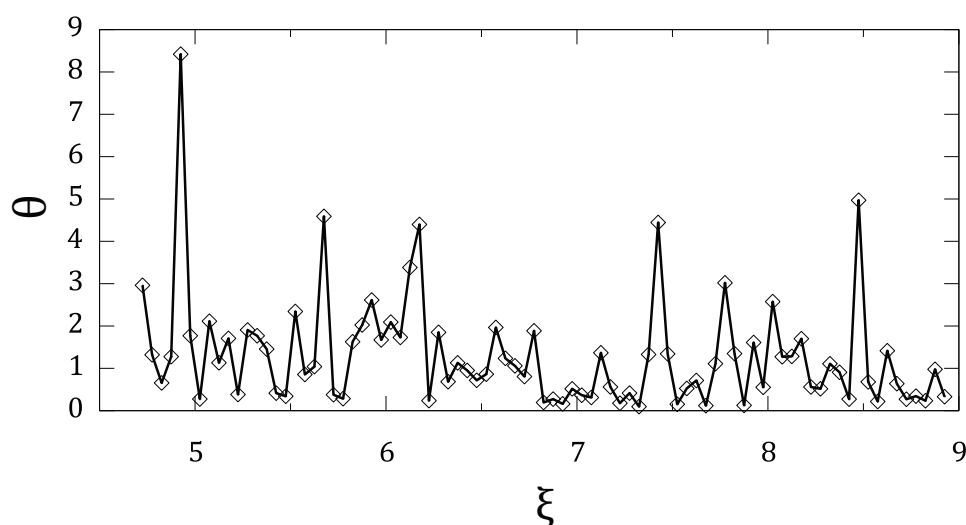


Figure 6.11 – Représentations des valeurs du coefficient d'incohérence θ le long du chemin de glissement avec la conformation interne de la DNM fixée.

l'incohérence entre les simulations adjacentes lorsque le ligand est placé entre la 7^{ème} et 9^{ème} paire de bases.

D'un point de vue général, cette fixation a permis de réduire partiellement l'influence des coordonnées réactionnelles orthogonales à ξ , tout en améliorant la cohérence de nombreux échantillonnages voisins. Cet impact reste cependant limité notamment pour des valeurs de $\xi > 7,5$ où la fixation du ligand semble avoir détérioré la qualité des échantillonnages en comparaison avec le chemin initial (Figure 6.8).

Finalement, afin de retourner vers un chemin plus réaliste, nous avons procédé à un dernier ensemble de simulations au cours duquel la conformation interne de la DNM a été relâchée.

6.4.4 Chemin final

A partir des structures avec la conformation interne de la DNM contrainte, chaque fenêtre le long de la coordonnée réactionnelle a été simulée 4 ns supplémentaires sans aucune contrainte sur les dièdres du ligand. Afin de calculer l'énergie libre du système, nous avons uniquement considéré les 2 dernières ns d'échantillonnages. Toutes les contraintes précédentes étant relâchées dès le début des simulations, les 2 premières ns d'échantillonnages sont considérées comme le temps nécessaire à l'équilibration des systèmes.

6.4.4.1 Profil d'énergie libre

Le relâchement des conformations internes du ligand engendrant une aug-

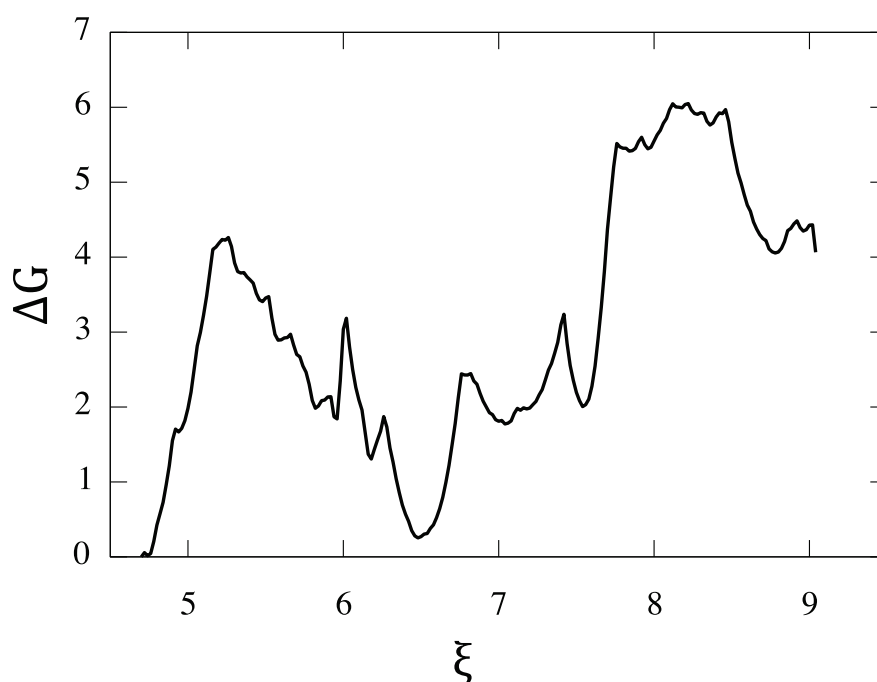


Figure 6.12 – Profil d'énergie libre (en kcal mol⁻¹) du glissement de la DNM le long du petit sillon de l'ADN, avec la conformation interne du ligand libéré.

mentation du nombre de degrés de libertés du système, il ne sera donc pas possible de comparer directement le PMF issu de ces simulations, à celui issu des simulations avec la structure du ligand fixée. Il sera cependant possible de comparer leurs barrières d'énergie. D'un point de vue général, ce nouveau PMF (Figure 6.12) montre des barrières d'énergie supérieures aux simulations précédentes, atteignant 6 kcal mol^{-1} pour le maximum d'énergie libre. Le puits d'énergie centré sur le site d'intercalation est quant à lui plus profond, ($\sim 6 \text{ kcal mol}^{-1}$ par rapport au maximum global, et $\sim 4 \text{ kcal mol}^{-1}$ par rapport au maximum local pour $\xi = 5,2$).

D'un point de vue entropique, l'augmentation du nombre de degrés de libertés devrait normalement contribuer favorablement à l'énergie du système, et donc réduire les barrières d'énergie libre le long du chemin. Cependant, dans notre cas, les barrières sont plus hautes. Ce phénomène peut s'expliquer par le fait que le relâchement de la conformation interne de la DNM, aura pour conséquence de réduire à nouveau la qualité de l'échantillonnage des degrés de liberté orthogonaux à la coordonnée réactionnelle ξ le long desquelles l'équilibration s'avère difficile.

Aux alentours des puits d'énergie, les conformations de la DNM sont bien adaptées, engendrant peu de différences énergétiques dans cette zone. Cependant, dans les zones de plus hautes énergies, rien ne nous permet d'affirmer que la conformation dans laquelle le ligand était précédemment fixé, est adaptée aux interactions avec l'ADN. De ce fait, l'augmentation des barrières d'énergie pourrait être la conséquence directe de transitions conformationnelles du ligand ainsi que de l'ADN qui devraient avoir lieu mais ne sont pas échantillonnées dans nos simulations.

6.4.4.2 Cohérence des échantillonnages

Le relâchement des dièdres du ligand a pour conséquence d'augmenter de façon dramatique l'incohérence entre des simulations voisines tout au long du chemin, et surtout au niveau des échantillonnages aux environs de la 6^{ème} paire de bases, zone correspondante à la barrière d'énergie de 4 kcal mol^{-1} que doit franchir

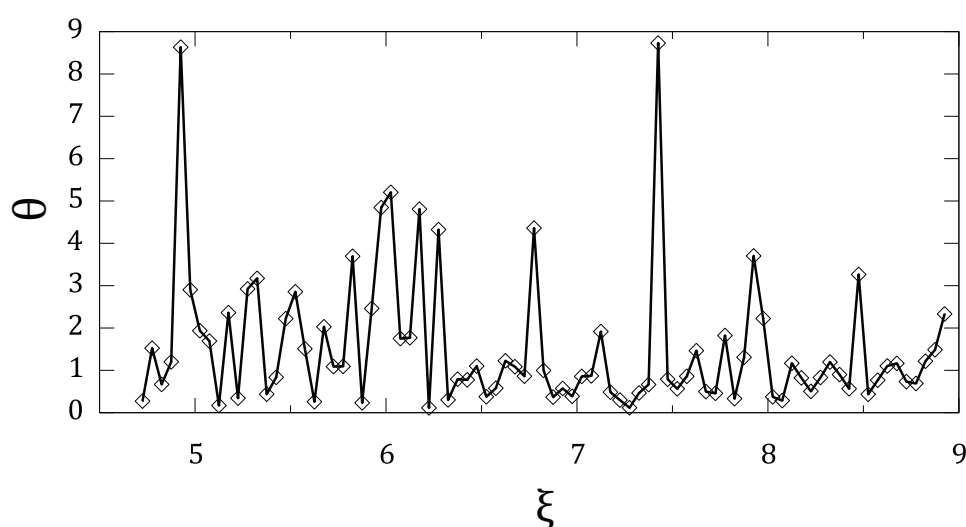


Figure 6.13 – Evolution du coefficient d'incohérence θ entre les fenêtres adjacentes le long du chemin de glissement après relâchement de la conformation interne de la DNM.

la DNM pour passer de la proximité de son site d'intercalation vers l'extrémité 5' du brin d'ADN.

Les analyses de cohérence que nous avons réalisées jusqu'à présent, nous ont permis de juger de la qualité de nos échantillonnages. Cependant, elles ne nous apportent aucune lumière quant au profil d'énergie libre en lui-même. De ce fait, il est indispensable de s'intéresser finalement à l'évolution structurale des systèmes le long du chemin de glissement issu des simulations avec la DNM relâchée.

6.4.4.3 Analyses conformationnelle

Les structures représentatives des fenêtres de simulations ayant servi à construire le chemin de glissement nous permettent en partie d'expliquer certains aspects du PMF.

Afin de faciliter la description du chemin, nous avons choisi 12 structures représentatives (Figure 6.15) issues de zones d'intérêt le long du PMF notées de **(a)** à **(l)** (voir Figure 6.14).

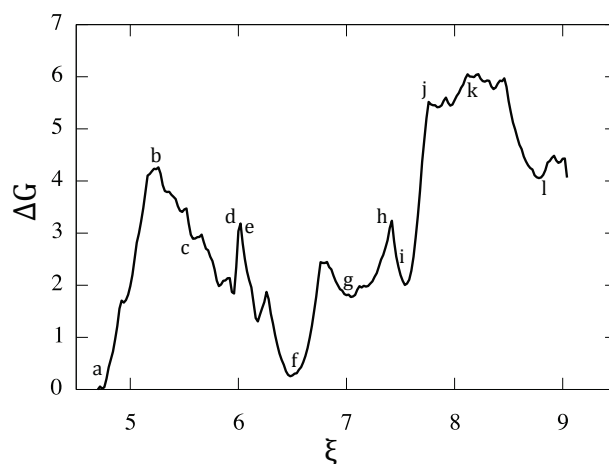


Figure 6.14 – PMF du chemin de glissement de la DNM le long du chemin de glissement. Les lettres a \rightarrow l correspondent aux structures représentées par la Figure 6.15.

Pour la structure **(a)** issue de la première fenêtre de simulation (caractérisée par un minimum local sur le PMF), la DNM est parfaitement placée dans le petit sillon de l'ADN, avec son cycle placé parallèlement au squelette de l'ADN. Une fois la barrière d'énergie passée, la structure représentative au niveau du maximum local **(b)** pour $\xi = 5,2$, est caractérisée par un petit sillon très ouvert, permettant au ligand de se réorienter dans le petit sillon.

Pour des valeurs de contraintes comprises entre 5,2 et 6, la DNM se place majoritairement de façon perpendiculaire au petit sillon de l'ADN comme illustré par **(c)**. Dans certaines fenêtres d'échantillonnage de cette zone, la DNM s'est totalement dissociée des bases de l'ADN tout en restant en contact avec le squelette de la double hélice.

Le pic d'énergie localisé au niveau de 6^{ème} paire de bases **(d)** est quant à lui caractérisé par le ligand partiellement intercalé en lieu et place de la 6^{ème} paire de bases qui s'est totalement ouverte vers le grand sillon.

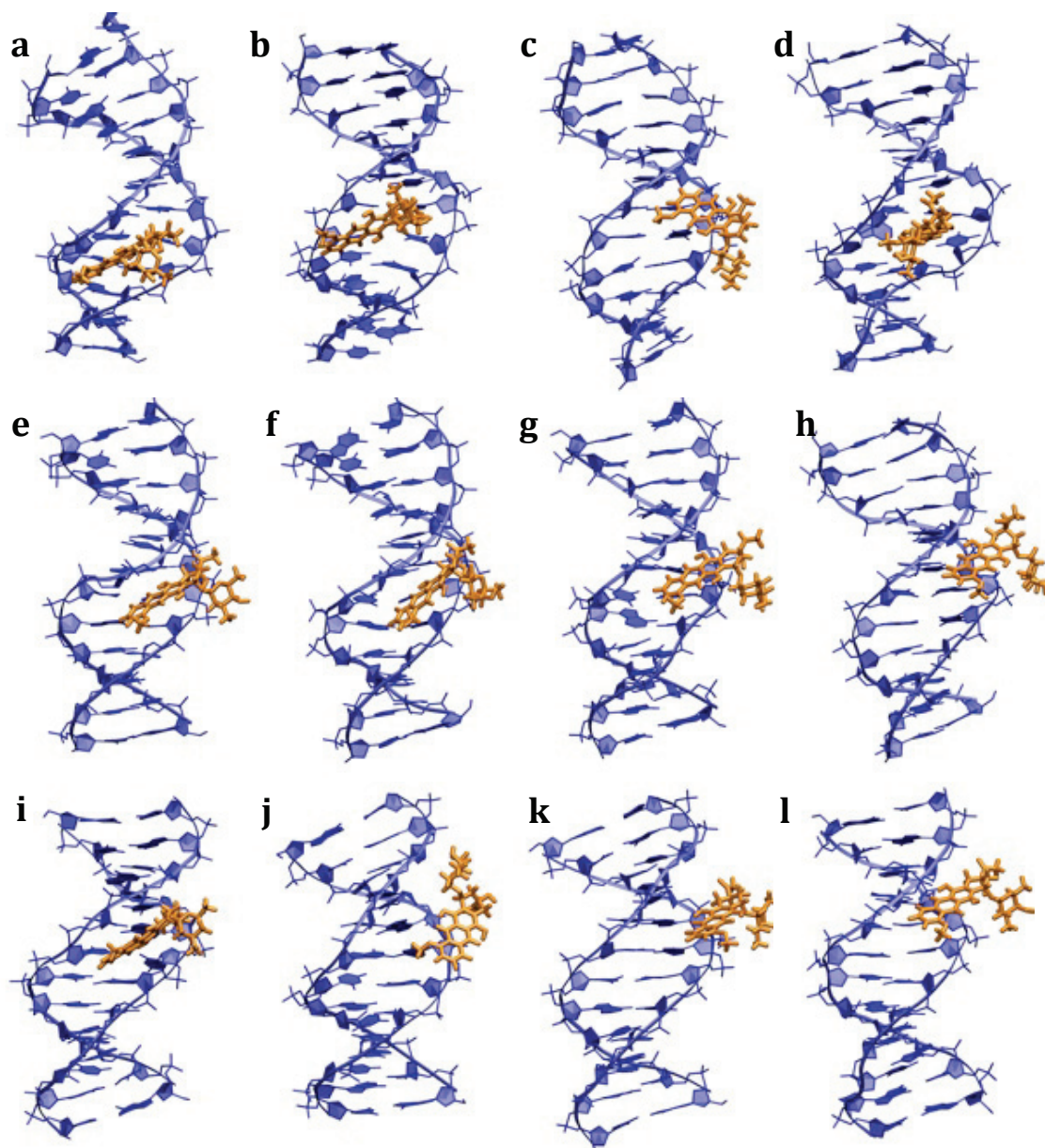


Figure 6.15 – Structures représentatives de fenêtres d'échantillonnage provenant de zones d'intérêt le long du chemin de glissement de la DNM dans le petit sillon de l'ADN.

Les structures suivantes montrent un remplacement du cœur rigide de la DNM dans le petit sillon (comme illustré par la structure **(e)**) jusqu'à atteindre le minimum local situé au niveau du site d'intercalation entre les paires de bases C6pG7 **(f)**.

Entre le minimum d'énergie et la structure **(i)**, la DNM alterne des phases où elle est parfaitement placée en profondeur dans le sillon avec des phases où elle est partiellement sortie comme illustré par les structures **(g)** et **(h)**.

Après avoir passé la seconde barrière d'énergie importante (4 kcal mol^{-1}) et atteint le plateau d'énergie à proximité de la huitième paire de bases, les structures représentatives nous montrent un ligand lié au petit sillon, mais très peu enfoui dans le sillon, comme s'il était prêt à se dissocier **(j)** et **(k)**). D'ailleurs, au niveau du plateau, nous avons pu constater que pour certaines fenêtres de simulation la DNM est complètement dissociée de l'ADN.

Finalement la DNM se replace plus profondément dans le sillon pour atteindre le minimum local à proximité de la 9^{ème} paire de bases de notre séquence (structure **(l)**).

Les structures représentatives le long du chemin de glissement nous permettent de compléter les analyses de cohérence d'échantillonnage effectuées précédemment. En effet, les deux zones de hautes énergies du PMF (soit pour $\xi \approx 5,2$ et $7,8 < \xi < 8,5$), sont caractérisées par des structures où la DNM est soit partiellement placée dans le sillon, soit parfois totalement dissociée de l'ADN. De telles structures verront ainsi leur entropie de translation/rotation fortement augmenter par rapport aux cas où le ligand est profondément enfoui dans le sillon. Or, la distance radiale entre le ligand et l'ADN est un degré de liberté orthogonal à la coordonnée de glissement contrainte, dont un échantillonnage exhaustif à chaque fenêtre est en théorie nécessaire. Néanmoins, et comme mentionné au Chapitre 5, l'augmentation de la liberté de rotation/translation du ligand aux grandes séparations rend l'équilibration le long de ce degré de liberté difficile et peut impacter la convergence des PMF présentés ici, ainsi que la cohérence entre fenêtres voisines, au même titre que les degrés de liberté internes de la DNM. Tous ces résultats nous permettent de conclure que les échantillonnages situés dans les zones de hautes énergies le long du chemin de glissements, ne sont pas convergés. Et ne le seront sans doute pas avec des temps de simulation raisonnables si nous partons des mêmes structures.

Si l'on s'intéresse à l'origine des déformations observées, il apparaît qu'au niveau des zones de hautes énergies observées sur le dernier PMF, de nombreuses fenêtres de simulation issues de l'échantillonnage originel ayant permis de construire notre chemin réactionnel, montrent des structures avec la DNM très peu enfouies dans le sillon, voire en position perpendiculaire comme dans le cas **(c)** (données non représentées). Par la suite, les nombreux échantillonnages additionnels que nous avons réalisés à partir de ces structures, n'ont pas permis de corriger ces déformations, conduisant même dans certains cas, à la dissociation du complexe.

6.4.5 Du glissement à l'intercalation

Cependant, malgré la mise en lumière de déformations influençant de façon importante le profil d'énergie libre, il semble important de noter la présence d'un puits d'énergie localisé au niveau du site d'intercalation.

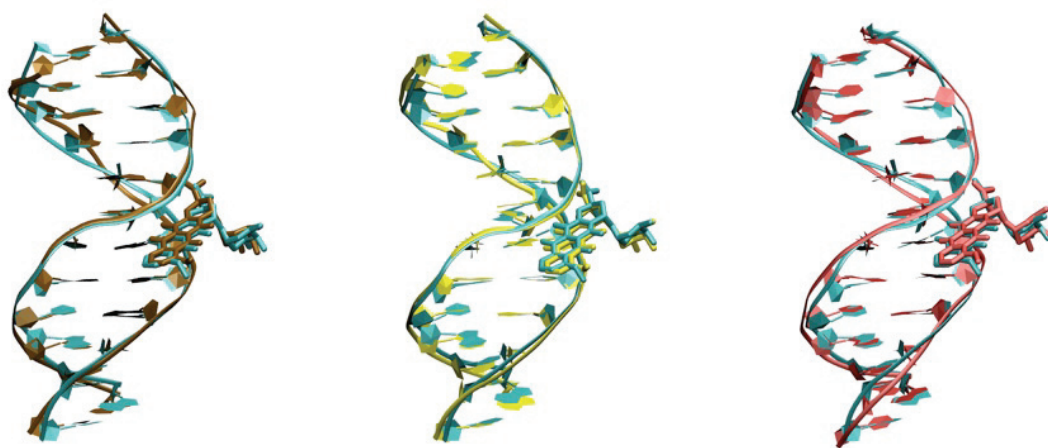


Figure 6.16 – Superposition de la structure MG1 (en bleu clair) avec la structure représentative du minimum d'énergie situé au niveau du site d'intercalation, provenant (i) du chemin initial (en marron à gauche), (ii) du chemin avec la structure de la DNM fixée (en jaune au centre), (iii) et du chemin après libération du ligand (en rouge à droite).

En effet, ce puits, en plus d'être présent sur les trois profils d'énergie libre que nous avons décrits précédemment, se trouve dans une zone où les échantillonnages semblent convergés (les coefficients d'incohérence étant très bas pour les trois chemins).

Lorsque l'on compare les structures représentatives de ce puits avec la structure MG1 décrite dans notre chapitre précédent, on peut observer qu'indépendamment du chemin de glissement, les structures se superposent parfaitement (voir Figure 6.16), aussi bien au niveau du ligand, qu'au niveau de la structure de l'ADN (Tableau 6.1).

Paramètres	MG1	Initial	Fixée	Relâchée
RMSD par rapport à MG1 (Å)	/	0,994	0,971	1,063
Rise C6pG7 (Å)	3,8	4	3,8	4
Roll C6pG7 (°)	-4	1,5	-1,4	2,2
Σ Roll A5-T8 (°)	-4	-10	-2	-2,4
Twist C6pG7 (°)	39	38	37	37
Σ Twist A5-T8 (°)	99	99	94	106
Largeur du petit sillon (Å)	4,9	4,8	4,7	4,8
Cycle A	Eq	Eq	Eq	Eq

Tableau 6.1 – Comparaison des paramètres de la structure de MG1 avec les paramètres des structures représentatives du minimum local situé au centre du pas C6pG7 le long des chemins de glissement initiaux, avec la DNM bloquée ainsi qu'avec la DNM relâchée.

Cette bonne superposition structurelle nous permet de pointer du doigt un potentiel processus de reconnaissance du site d'intercalation par le ligand : lors du déplacement de la DNM le long du petit sillon, la présence de ce puits énergétique va permettre de stabiliser le ligand au niveau de son site d'interaction dans une

conformation correspondant à MG1. A partir de cette position, le ligand pourra soit continuer à glisser vers la prochaine paire de bases soit entamer son intercalation entre les paires de bases via le mécanisme que nous avons décrit dans le chapitre précédent.

6.5 Discussion et conclusions

La contrainte que nous avons développée pour ces travaux nous permet de simuler le glissement d'une molécule le long du petit sillon de l'ADN. Contrairement aux contraintes classiques utilisées dans des travaux précédents [349] (basée notamment sur les distances interatomiques), cette contrainte nous permet de définir la position du ligand dans le petit sillon indépendamment des potentielles déformations de la séquence d'ADN. Après des tests de validation, nous avons pu utiliser cet outil pour créer le chemin de glissement de la DNM le long du petit sillon de l'ADN.

Après des premiers résultats montrant des barrières d'énergie libre surprenamment hautes (~ 16 kcal mol⁻¹), en totale contradiction avec un potentiel glissement le long du sillon, nous avons pu améliorer nos échantillonnages en réduisant puis en libérant la conformation interne de la DNM pour atteindre des énergies maximales de l'ordre de ~ 6 kcal mol⁻¹.

Cependant, les études structurelles et de cohérence d'échantillonnage basées sur les travaux de Zhu et Hummer [305], nous obligent à conclure que certaines de nos fenêtres de simulations (notamment dans les zones de hautes énergies) sont loin d'être convergées, sans doute en conséquence d'un mauvais échantillonnage initial de ces zones.

Malgré cela, les premiers résultats que nous avons obtenus sont plus qu'encourageants. En effet, nos simulations montrent des énergies de glissement inférieures à celle obtenues par Furini et *al.* [348] (de l'ordre de 9 kcal mol⁻¹), et équivalentes à celle de Vargiu et *al.* [349] pour un système proche du nôtre.

Au vu de nos résultats, il nous semble intéressant de poursuivre notre étude en recréant un chemin partant du point de départ tout en améliorant la qualité des échantillonnages initiaux afin de mieux contrôler les déformations du complexe. Cependant, une telle solution pourrait s'avérer très coûteuse en temps de calcul au regard de la complexité du système.

Les simulations au cours desquelles nous avons bloqué la conformation interne du ligand ayant permis de nettement améliorer la convergence des profils d'énergie libre, une alternative logique pourrait consister à construire un nouveau chemin en fixant ces conformations, puis de les relâcher progressivement. De telles simulations sont en cours de réalisation dans notre laboratoire.

Finalement cette étude nous a permis de mettre en lumière le rôle important du glissement de la DNM le long du petit sillon dans la localisation de son site d'interaction. En effet, nos premiers résultats montrent que le passage de la DNM à proximité du site d'intercalation va stabiliser le complexe dans une conformation que nous avons identifiée lors de notre étude précédente comme un intermédiaire important du mécanisme d'intercalation.

Ces observations (notamment la présence du puits énergétique) devront cependant être confirmées au cours de nos prochaines simulations.

Conclusions Générale

Depuis des décennies, les molécules intercalantes de l'ADN sont reconnues et utilisées pour leur activité thérapeutique. En dépit de cela, et malgré un très grand nombre d'études traitant de ce type de molécules dans de nombreux domaines (plus de 11 000 publications traitant d'intercalants de l'ADN étant répertoriées à ce jour sur PubMed), la compréhension de leur mécanisme d'action ressemble à un puzzle dont de nombreuses pièces restent manquantes. Les études expérimentales consacrées à ce sujet ont toutefois permis d'appréhender l'impact structural de l'intercalation d'une molécule entre deux paires de bases, ainsi que de se positionner vers un processus d'intercalation complexe impliquant plusieurs étapes (le nombre de ces étapes étant sujet à débat). Cependant les détails du chemin menant une molécule intercalante d'un état libre en solution à son site d'intercalation reste une énigme, qu'aucune étude expérimentale (que ce soit *in vivo* ou *in vitro*) ne semble pouvoir résoudre actuellement. A l'heure où les supercalculateurs et les modèles physiques nous permettent de simuler des systèmes macromoléculaires de plus en plus gros, avec des représentations de plus en plus réalistes et dans des temps de calcul de plus en plus limités, les simulations de dynamiques moléculaires apparaissent comme une alternative de choix dans l'étude de tels systèmes.

Au cours de cette thèse, j'ai donc tenté d'améliorer notre vision des mécanismes pouvant guider et/ou influencer une molécule intercalante (une daunomycine), le long du chemin la menant à son site d'intercalation, via des simulations de dynamique moléculaire.

La daunomycine, au même titre que l'ADN est un objet flexible (même si cette flexibilité reste limitée à son cycle A et son sucre daunosamine). La première partie de mes travaux a permis de montrer que le cycle flexible de la daunomycine présente deux conformations principales de stabilités quasi-équivalentes. En présence

d'ADN, cet équilibre est fortement perturbé, conséquence directe de nombreuses interactions interfaciales, notamment un pont salin très stable entre l'ammonium du ligand et les phosphates de l'ADN. Dans le cas de simulations en conditions minimales cette liaison sera surstabilisée avec un effet dramatique sur la structure du complexe. D'une manière plus générale, ce résultat nous permet de pointer la nécessité de porter une grande attention aux conditions ioniques utilisées, en fonction du système simulé. En conséquences, toutes les simulations suivantes ont été réalisées en condition physiologique. Finalement, cette étude nous a permis de mettre en évidence, dans le cas où la daunomycine se trouve liée au petit sillon, une influence de la configuration du ligand sur le roll entre deux paires de bases successives de l'ADN, pouvant potentiellement jouer un rôle dans l'initiation de l'intercalation.

Une étude de dynamique moléculaire datant de 2008 utilisant une représentation tous atomes d'un complexe ADN/daunomycine a permis de mettre en lumière un processus d'intercalation impliquant deux états (un état lié au petit sillon de l'ADN ainsi qu'un état intercalé), sans pour autant parvenir à caractériser un chemin réactionnel les reliant. Cette lacune mise en évidence, la suite logique de mes travaux était toute trouvée. A partir d'une contrainte permettant de biaiser le système au minimum, nous avons pu créer par umbrella sampling, un chemin réactionnel menant la daunomycine d'un état lié au petit sillon à son site d'intercalation. Ces étapes montrent des énergies d'activation en accord raisonnable avec les valeurs expérimentales. Cependant au cours de cette étude nous n'avons porté que très peu d'attention à l'influence du milieu environnant, composé d'eau et d'ions. L'eau est depuis longtemps reconnue comme un partenaire à part entière dans les processus de formation des complexes ADN/ligand. Une suite logique à ces travaux, pourrait viser à caractériser le rôle exact de l'eau dans le mécanisme d'intercalation, et notamment dans la stabilisation d'intermédiaires structuraux essentiels. De même, au vu de l'influence non négligeable de la concentration saline que nous avons pu mettre en évidence, il serait important d'approfondir les recherches dans cette direction pour le futur.

Le point le plus intéressant de ce chemin correspond à la présence d'un état intermédiaire jamais décrit auparavant pour la daunomycine. Cet état présente une conformation déjà observée dans les structures de complexes ADN/protéines intercalantes de l'ADN. Se pose cependant la question de savoir si un tel état est caractéristique de toutes les molécules intercalantes ou si il est propre au mécanisme de la daunomycine, voire des autres molécules de la famille des anthracycline. Dans ce sens, il serait intéressant de poursuivre ces travaux en caractérisant le processus d'intercalation de molécules proches structurellement parlant de la daunomycine dans un premier temps (telle l'adriamycine par exemple) puis, dans un second temps, à partir de molécules bien plus complexes, voir même des protéines.

Malgré l'amélioration de la description du chemin menant la daunomycine à son état intercalé, une question importante reste sans réponse : comment le ligand fait-il pour reconnaître spécifiquement son site d'intercalation ? Est-ce par une diffusion 3D dans le milieu lui permettant de trouver au « hasard » sa cible, ou bien après une phase de liaison aspécifique à l'ADN le ligand « glisse » le long du petit sillon jusqu'à son site d'interaction ? Tenter de répondre à cette question en recréant ce glissement et ainsi voir si ce type de mouvement s'avère réaliste ou non pour la daunomycine, a été le sujet de la dernière partie de ma thèse. Développer une contrainte permettant à un ligand de glisser le long du petit sillon de l'ADN en lui laissant un maximum de liberté, s'est avéré beaucoup plus compliqué que ce à quoi nous nous attendions. Notamment vis à vis des déformations de la double hélice. Nous sommes finalement parvenus à développer une contrainte avec laquelle nous avons pu déplacer la daunomycine dans le petit sillon. Nos premiers résultats ont abouti à des énergies écartant toutes possibilités de glissement du ligand. Leurs analyses ont finalement mis en évidence un manque de convergence de nos échantillonnages, et ce, dès la phase de construction du chemin. Malgré de mauvais échantillonnages, les derniers résultats que nous avons présentés sont très encourageants au vu des énergies obtenues. Finalement, cette première utilisation de la contrainte nous a permis de mettre en évidence une zone de stabilisation le long du chemin de glissement située exactement au niveau du site

d'intercalation de la daunomycine. Les structures issues de cette zone, s'avèrent très similaires à l'état lié au petit sillon caractérisé dans l'étude précédente.

Cette similarité nous permet d'envisager un processus global d'intercalation de la daunomycine impliquant une phase de glissement/localisation suivie de l'intercalation en elle-même. Des simulations avec des degrés de liberté réduits (permettant d'améliorer la convergence des simulations), sont actuellement en cours de réalisation. La caractérisation d'un tel processus permettrait d'augmenter notre compréhension des événements guidant une molécule vers sa cible, et d'ainsi potentiellement développer des molécules ciblant l'ADN de façon plus spécifique. La caractérisation du glissement de la daunomycine n'est qu'un point de départ. Afin d'obtenir une vision d'ensemble du processus de reconnaissance moléculaire, il sera nécessaire de caractériser le glissement de nombreuses autres molécules. Dans ce sens, la méthode que nous avons développée nous permet de raisonnablement envisager l'étude du glissement à partir de séquences d'ADN plus étendues, tout en augmentant progressivement la complexité des ligands, pour à terme, aboutir à la caractérisation du glissement de protéines le long de l'ADN.

ANNEXE A

Evolution des paramètres hélicoïdaux

Résumé des paramètres hélicoïdaux des différents pas d'un ADN-B canonique de séquence d(GCGCACGTGCGC)₂ après 50 ns de simulations en conditions minimales (bleu) et physiologiques (rouge) de concentrations en sels. *ps* correspondant à petit sillon, et *GS* à grand sillon.

Figure A.1 – Pas C2pG3

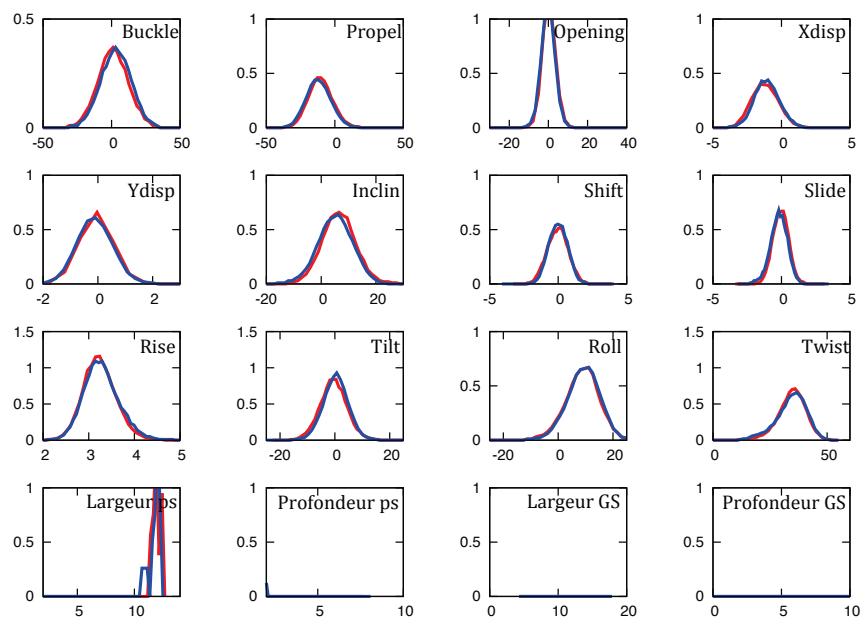


Figure A.2 – Pas G3pC4

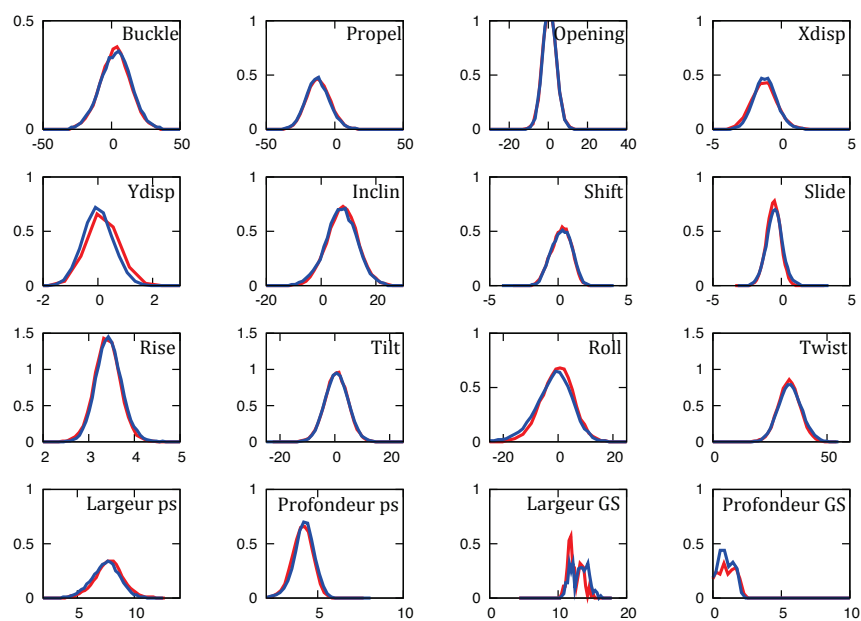


Figure A.3 – Pas G4pA5

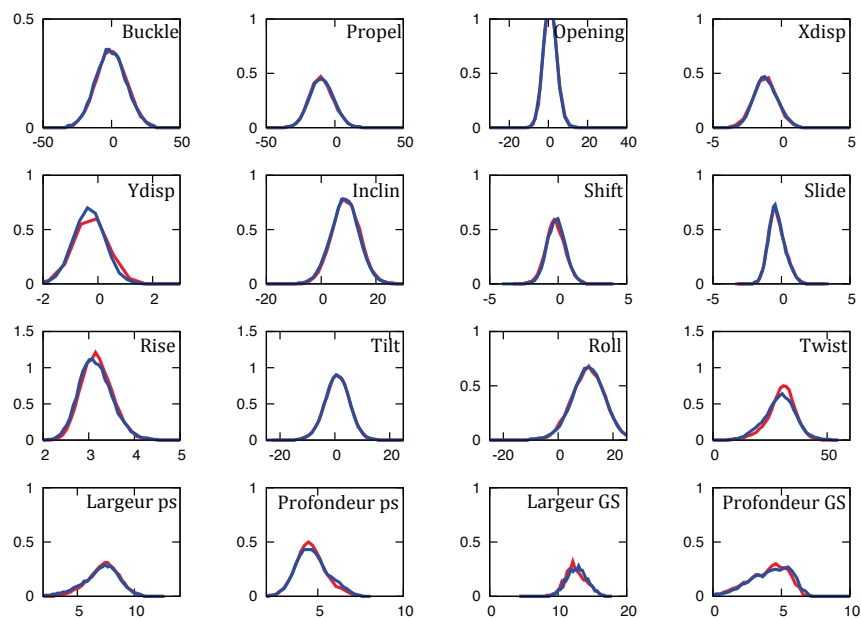


Figure A.4 – Pas A5pC6

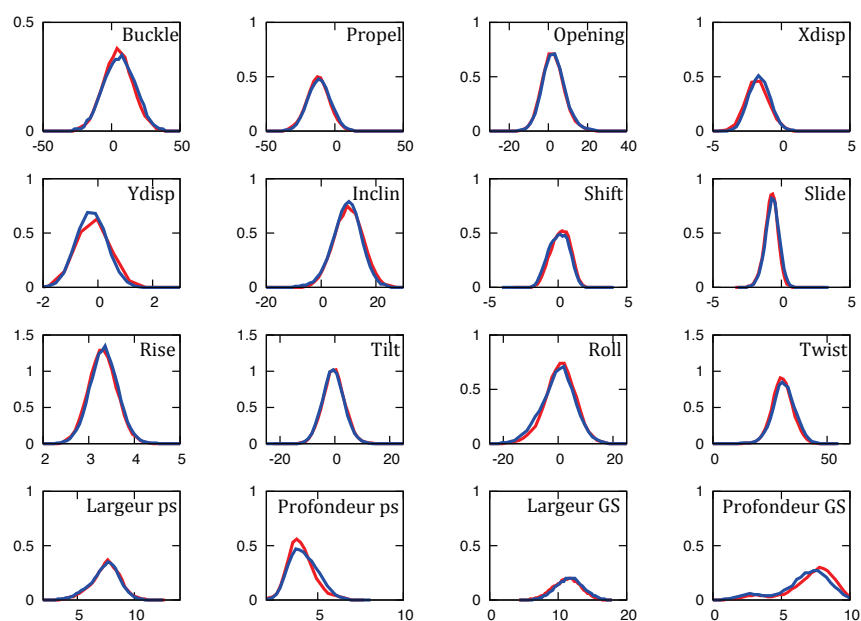


Figure A.5 –Pas C6pG7

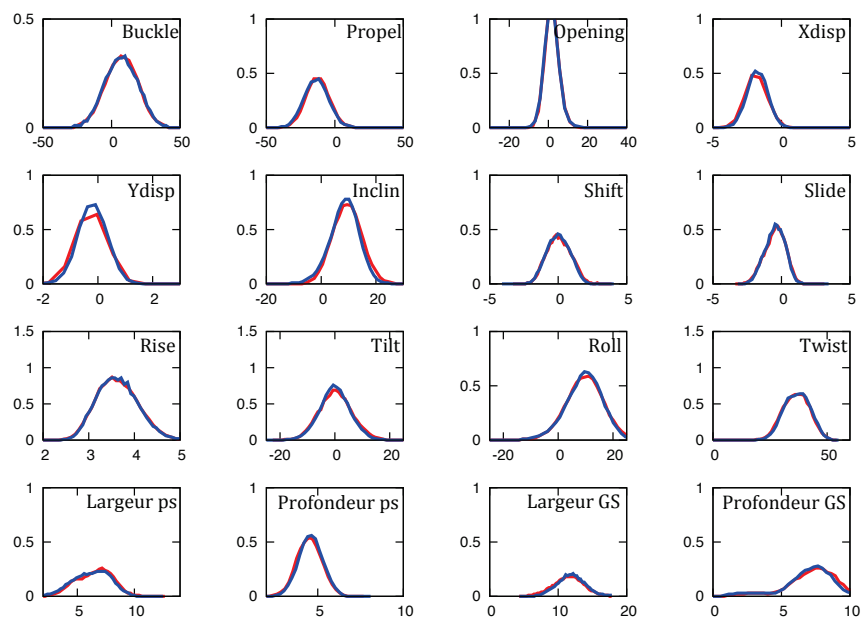


Figure A.6 – Pas G7pT8

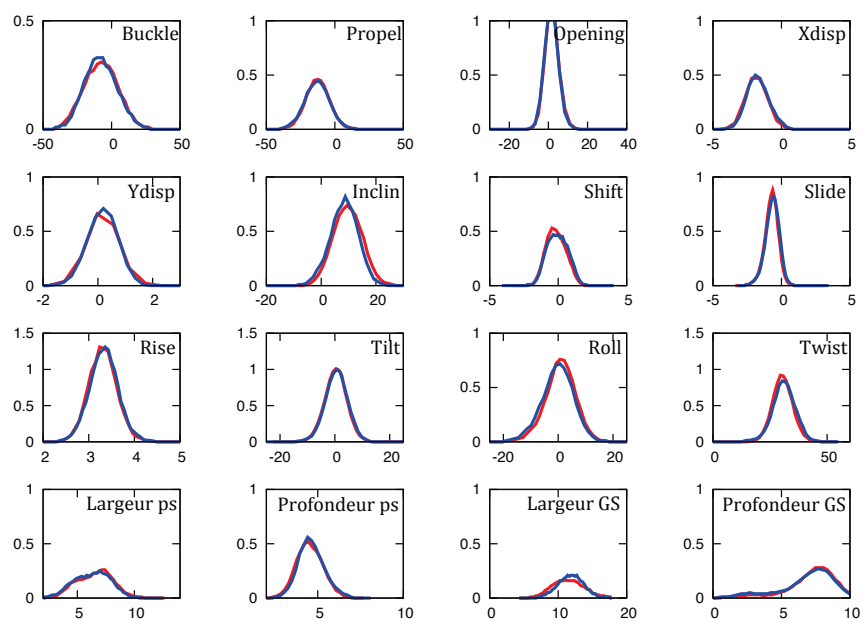


Figure A.7 – Pas T8pG9

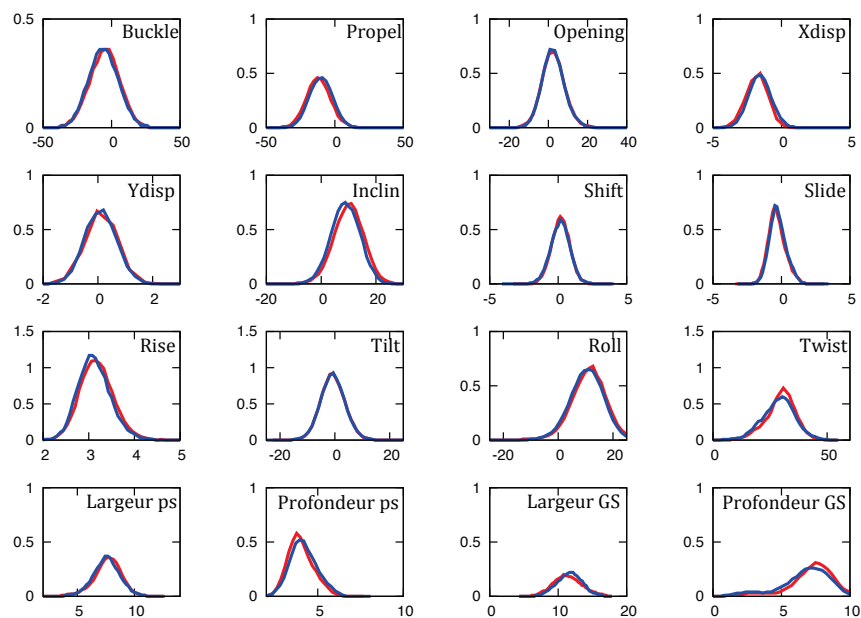


Figure A.8 – Pas G9pC10

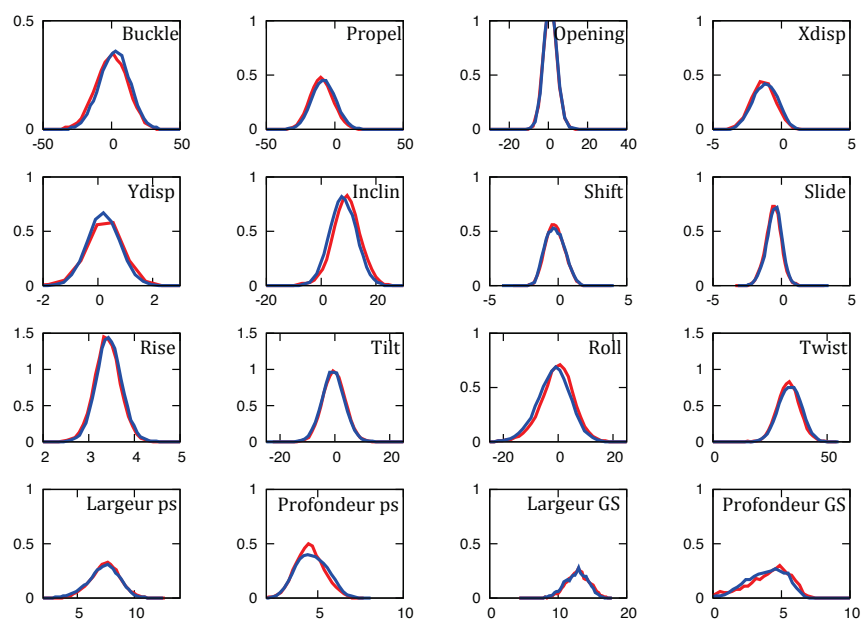
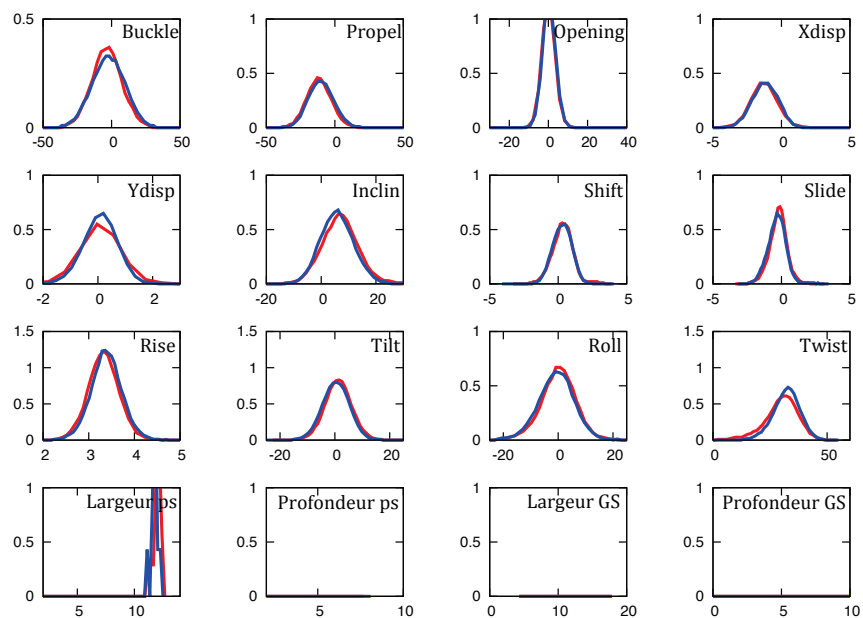


Figure A.9 – Pas C10pG11



ANNEXE B

Article publié

Multistep Drug Intercalation: Molecular Dynamics and Free Energy Studies of the Binding of Daunomycin to DNA

Matthieu Wilhelm,[†] Arnab Mukherjee,^{*,‡} Benjamin Bouvier,[†] Krystyna Zakrzewska,[†] James T. Hynes,^{*,§,||} and Richard Lavery^{*,†}

[†]Bioinformatics: Structures and Interactions, Bases Moléculaires et Structurales des Systèmes Infectieux, Univ. Lyon I/CNRS UMR 5086, IBCP, 7 Passage du Vercors, Lyon 69367, France

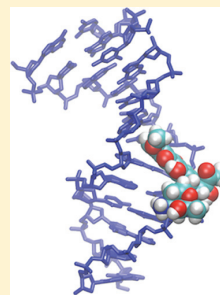
[‡]Chemistry Department, Indian Institute of Science Education and Research, Pune, 411021, India

[§]Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215, United States

^{||}Chemistry Department, Ecole Normale Supérieure, CNRS UMR 8640, 24 rue Lhomond, 75005 Paris, France

Supporting Information

ABSTRACT: Atomic-scale molecular dynamics and free energy calculations in explicit aqueous solvent are used to study the complex mechanism by which a molecule can intercalate between successive base pairs of the DNA double helix. We have analyzed the intercalation pathway for the anticancer drug daunomycin using two different methods: metadynamics and umbrella sampling. The resulting free energy pathways are found to be consistent with one another and point, within an equilibrium free energy context, to a three-step process. Daunomycin initially binds in the minor groove of DNA. An activated step then leads to rotation of the drug, coupled with DNA deformation that opens a wedge between the base pairs, bends DNA toward the major groove, and forms a metastable intermediate that resembles structures seen within the interfaces between DNA and minor-groove-binding proteins. Finally, crossing a small free energy barrier leads to further rotation of daunomycin and full intercalation of the drug, reestablishing stacking with the flanking base pairs and straightening the double helix.



■ INTRODUCTION

An important class of pharmacologically active molecules function by intercalation between the base pairs of DNA, subsequently interfering with DNA function or replication, often by blocking the action of topoisomerases, either by stabilizing the cleaved DNA produced by topoisomerase II¹ or by preventing topoisomerase I from binding to DNA.² Some intercalators can also act as dual topoisomerase I/II inhibitors.^{3,4} Topoisomerase inhibition subsequently blocks DNA replication and leads to cell death. Intercalators are characterized by planar, conjugated, fused ring systems, whose dimensions resemble DNA base pairs. Insertion between consecutive base pairs (BPs) shields the hydrophobic rings of the intercalator from water, while favorable π -stacking is maintained within the double helix.^{5,6} Despite the medical applications of intercalators as anticancer drugs,^{7,8} and many years of study, a detailed mechanism of the intercalation process has not yet been determined.^{9,10} As structural studies have shown, intercalation must clearly lengthen DNA, doubling the inter-BP rise at the intercalation site (from approximately 3.4 to 6.8 Å) and generally reducing twist, but the pathway (or pathways) leading to this state from the free intercalator in solution is unknown. Kinetic studies point to a complex multistep process.^{11–19} Notably, for the final intercalation step, current evidence suggests that the intercalator is involved in creating the intercalation site, rather than passively waiting for such a site to occur spontaneously,^{20–22} despite the fact that

significant fluctuations in DNA rise are known to occur at room temperature.²³

Here, we use atomic-resolution molecular dynamics in aqueous solution, coupled with two different free energy evaluation procedures, to study the details of the intercalation process for daunomycin (also known as daunorubicin),^{24,25} an anthracycline anticancer drug⁸ that binds to DNA and blocks its replication. The structure of DNA–daunomycin complexes has been resolved by crystallography^{26–29} and by NMR spectroscopy^{30,31} and shows the characteristic features of most intercalating drugs. Kinetic studies of daunomycin by Chaires et al.¹⁶ suggested a three-step model for intercalation, involving an initial “outside” binding to DNA and a final “reshuffling” of the drug within the intercalation site. Other studies by Rizzo et al. suggested a five-step process.¹⁹ Using detailed simulations, we hope to be able to clarify the structural and energetic features of the intercalation pathway and to identify any intermediate states, providing information that cannot easily be obtained with current experimental techniques.

This work builds on our previous umbrella-sampling studies of daunomycin binding to DNA that revealed a stable, minor-groove-bound state reached by a barrierless route from the free intercalator state.³² This state did not require significant DNA deformation and could explain the outside bound step

Received: February 19, 2012

postulated by Chaires and co-workers.¹⁶ Our former study also indicated that the intercalator was involved in creating the intercalation site and enabled us to identify some key motions of the DNA–daunomycin–water system (e.g., DNA rise, roll, and bend variations) involved in the transition from the groove-bound to the intercalated state. Although this study provided an estimate of the free energy barrier, the methodology employed could not provide a detailed description of the transition. The present study fills this gap.

Since intercalation is a conformationally complex process, involving significant DNA deformation coupled with a major repositioning of the drug molecule, we have derived optimal free energy pathways with two distinct methods. The first uses nonequilibrium, metadynamics simulations³³ in a specifically designed collective variable space. Metadynamics enables a system to sample large volumes of conformational space in a reasonable computational time by regularly adding local destabilizing potentials to the natural energy hypersurface. It has already been successfully used to study a variety of binding and unbinding processes.^{34–36} The second method, in common with our earlier studies,³² uses equilibrium umbrella sampling with the aid of a biasing potential to drive the system between the desired states in regular steps. In order to study the separation of two molecules using this technique, we have recently developed a new restraint that avoids having to specify a separation pathway. This approach has already been successfully used to study the mechanism of protein–DNA unbinding.^{37,38} The results of these two independent approaches are found to be consistent and reveal a passage from the groove to the intercalation site involving a metastable intermediate and two transition states. In the metastable intermediate state, daunomycin is partially intercalated at a 5'-CpG-3' step, opening up a wedge angle between the two base pairs on the minor-groove side and bending DNA toward the major groove. This structure is similar to those seen at the interface between minor-groove-binding proteins and DNA,³⁹ where hydrophobic amino acid side chains are also found to partially intercalate into the double helix.

METHODOLOGY

Simulation Setup. Following the procedure set out in our earlier study,³² the initial conformation of daunomycin and of its DNA intercalation site were drawn from crystallographic results,²⁸ using the PDB structure 1D11. Since this structure contains two daunomycin molecules, we extracted a single intercalation site at a CpG step, including one daunomycin, contained within a three nucleotide pair 5'-ACG-3' segment. The 5'-AT base pair was retained because the charged sugar moiety of daunomycin (see Figure 1) lies in the minor

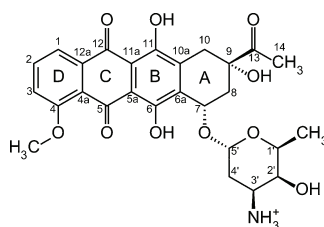


Figure 1. Chemical structure and notation of daunomycin. The rings B, C, and D constitute the planar, conjugated anthraquinone moiety that intercalates between adjacent DNA base pairs. The A-ring is nonplanar and flexible and carries the cationic daunosamine sugar at position C7.

groove, close to this base pair. The ACG segment was then extended using JUMNA,⁴⁰ by adding canonical B-DNA arms on either side to form a double-stranded dodecamer with the sequence d-(GCGCACGTGCGC)₂, where the underline indicates the position of the central C6pG7 intercalation site. Note that a CpG step (that is, a CG base pair followed by a GC base pair in the 5'–3' sense) has been shown to be the strongest binding site for daunomycin.^{41–43} Note also that the dodecamer has a palindromic base sequence. This implies that the orientations of the drug created by 180° rotations around the C5–C12 axis of the anthraquinone segment (see Figure 1), either in the intercalation site or centrally positioned in a DNA groove, will be symmetrically equivalent.

Simulation Protocol. DNA was simulated using the AMBER parm99/bsc0 force field^{44,45} that corrects earlier problems connected with unusual backbone states and leads to conformational distributions in good agreement with experiment.^{45,46} Daunomycin was represented using the general AMBER force field (GAFF).⁴⁷ Its atomic charges were calculated with the Merz–Kollman electrostatic potential potential-fitting procedure⁴⁸ in the Gaussian quantum chemistry package,⁴⁹ using a Hartee-Fock wave function obtained with a 6-31G* basis set for compatibility with the partial charges used for DNA in the AMBER force field (for further details see the Supporting Information).

The initial conformation of the daunomycin–DNA complex was simulated in a cubic box with periodic boundary conditions. The solute was surrounded by roughly 9000 TIP3P water molecules,⁵⁰ leaving at least 10 Å space between the solute and the faces of the box. The solute charges were compensated in two possible ways, either by adding a minimal number of ions to neutralize the system (22 Na⁺ and 1 Cl[−]) or by adding sufficient additional salt (NaCl) to reach a concentration of 0.15 M, close to physiological conditions. In both cases, ion parameters from Dang et al. were used.⁵¹ The system was equilibrated using energy minimization followed by careful heating to 300 K with harmonic restraints on the solute atom positions. These restraints were relaxed in a series of energy minimizations, followed by short molecular dynamics simulations. The system was then equilibrated for 5 ns, with no positional restraints, and using an isothermal–isobaric ensemble. Subsequent production simulations were carried out using either the GROMACS⁵² or NAMD⁵³ program suites. All simulations used a 2 fs time step. For the metadynamics simulations, iterative restraints were applied to all bond lengths, while for the umbrella-sampling dynamics simulations only X–H bond lengths (where X is any heavy atom) were restrained.

Metadynamics Free Energy Pathways. The first pathway generation approach uses metadynamics, which is a nonequilibrium sampling method, carried out using chosen collective variables.^{33,54} In order to generate a pathway between the intercalated and groove-bound states of daunomycin, we use three collective variables: *X*, which describes the distance of the drug from DNA; *Y*, which describes its movement in the direction of the DNA helical axis; and *θ*, which describes the orientation of the planar anthraquinone moiety (which is required to describe the reorientation of the drug between the groove-bound and intercalated states). *X* and *θ* are the same variables used in our earlier study.³² The geometrical definition of these variables is given in the Supporting Information. We also follow the procedure of this publication in examining the intercalation process in the reverse sense. In order to make the drug move away from the intercalation site as the simulation progresses, destabilizing Gaussian potentials are added to the region of collective variable space currently being sampled, thus forcing the system to explore higher energy regions. When convergence is reached, the Gaussian potentials allow the underlying free energy surface to be reconstructed. Metadynamics was carried out with GROMACS using the GROMETA implementation.⁵⁵ Gaussian potentials were added every 0.2 ps with a height of 0.024 kcal mol^{−1} and widths of 0.1 Å in the *X* and *Y* coordinates and of 0.2 radians in the *θ* coordinate. Several independent simulations were carried out in 1D (using the *X* coordinate), in 2D (using the *X* and *θ* coordinates), and in 3D (using the *X*, *θ*, and *Y* coordinates). We present here only the 3D simulations, that required of the order of 100 ns of sampling.

Umbrella Sampling Free Energy Pathways. The second pathway generation approach uses umbrella sampling. This is an equilibrium technique that constrains the system at a series of intermediate steps along a chosen coordinate using a biasing potential. Providing the sampling at successive steps overlaps, the impact of the biasing potential can be removed using the weighted-histogram analysis method (WHAM).⁵⁶ In order to separate daunomycin from DNA without imposing a specific spatial coordinate, we replaced the specific coordinates used in our earlier work³² with a more general approach developed for studying macromolecular dissociation processes.³⁷ This involves a restraint on the minimal approach distance between any pair of non-hydrogen atoms at the interface between two molecules. Any atom pair falling below the chosen distance limit leads to a quadratic penalty energy and consequently to atomic forces pushing the pair apart to the desired minimal distance. To avoid the two molecules drifting apart once they no longer interact strongly, we also impose a similar quadratic penalty on the closest atom pair at the interface, if this pair lies beyond the desired distance (see Supporting Information for further details). By increasing the restraint distance in small steps (0.15 Å) and adequately sampling the corresponding conformational fluctuations of the system (1–2 ns sampling before changing the restraint distance, followed by 2–4 ns of further simulation) we can build up an accurate view of the free energy pathway.

We also used umbrella sampling to analyze the internal flexibility of daunomycin and, in particular, of its flexible A-ring. This was sampled using the C7–C8–C9–C10 dihedral angle (hereafter termed τ) to scan the possible conformations of this ring. The dihedral was changed in 5° steps and restrained with a force constant of 0.15 kcal mol⁻¹ deg⁻¹. We performed 1 ns of sampling before each change. Each point was then sampled for a further 3 ns. The 30 sampling windows along each pathway required 120 ns of simulation. The corresponding free energy curves were again obtained using the WHAM procedure.⁵⁶

Conformational Analysis. A variety of tools were used for conformational analysis. The overall characteristics of molecular dynamics trajectories were analyzed using structures (“snapshots”) extracted at regular 1 ps intervals. These structures were superposed⁵⁷ and compared using root-mean-square difference (rmsd) calculations restricted to non-hydrogen atoms. Representative conformations at given points along the free energy pathways were chosen as the structure closest to the average structure of the snapshots sampled at this point, again using rmsd calculations. DNA structure was analyzed using the Curves+ program,⁵⁸ which was extended for the present study by including the possibility of locating a DNA-bound ligand. This enables a reference frame on the ligand to be located with respect to a reference frame on the DNA helical axis. Here, the position of daunomycin with respect to the center of the DNA helical axis is simplified to two parameters, the overall translation (L-Tra) and the overall rotation (L-Rot) between the two frames (see Supporting Information for details).

All molecular representations in this work were prepared using VMD,⁵⁹ and 2D and 3D graphics were prepared with Gnuplot, Mathematica, and MatLab.

RESULTS

Metadynamics Intercalation Pathway. Following a 110 ns metadynamics trajectory in the three collective variables described above (X , θ , Y) and using minimal salt conditions, we extracted a 2D free energy surface by averaging over the values along the Y coordinate. The results in Figure 2 show several local free energy minima that we have labeled IC, IM, MG1, MG2, and MG3. A dotted line indicates the lowest free energy path between these minima and indicates the associated transition states, labeled TS1–TS4. IC corresponds to the intercalated state of daunomycin, while MG1, MG2, and MG3 all have the drug lying within the minor groove. MG1 is the deepest of these minima, and we will not discuss the other groove-bound minima further, since drug movement within the

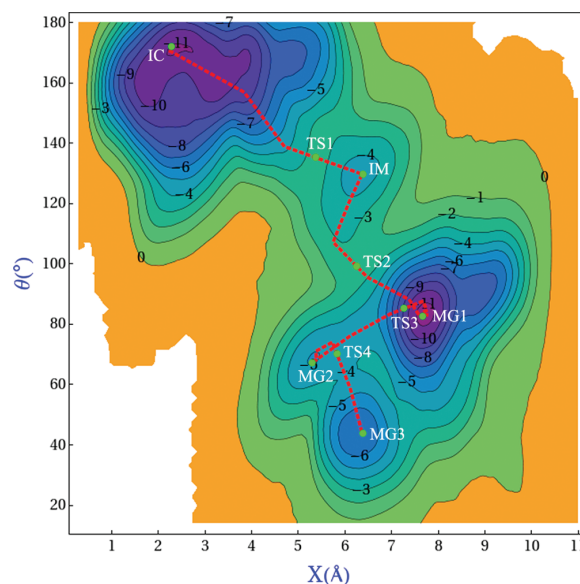


Figure 2. 2D contour plot of the metadynamics free energy trajectory obtained using three collective variables (X , θ , and Y). Values corresponding to displacements along the Y -axis (roughly aligned with the DNA helical axis) have been Boltzmann-averaged. Free energy contours are drawn at 1 kcal mol⁻¹ intervals and colored from yellow to purple as the system becomes more stable. The lowest free energy path through the most important local minima and the corresponding transition states is shown.

minor groove is still under investigation and is not our focus here.

The most striking feature of Figure 2 is a well-defined free energy basin corresponding to an intermediate state IM, lying between the highest transition states. This minimum is reached after daunomycin has moved roughly 4.5 Å along X , away from the intercalated state IC, and has rotated by roughly 40° in the θ coordinate. As shown in Table 1, IM lies 7.3 kcal mol⁻¹ above

Table 1. Free Energy Differences (kcal mol⁻¹) with Respect to the Intercalated State along the Metadynamics and Umbrella-Sampling Paths toward the Minor-Groove-Bound State

pathway	IC	TS1	IM	TS2	MG1
metadynamics	0	9.9	7.3	9.7	0.4
umbrella sampling	0	9.0	7.7	9.0	2.7

IC and 6.9 kcal mol⁻¹ above the best minor-groove-bound state MG1. The intermediate state is flanked by two transition states, TS1, 2.6 kcal mol⁻¹ above IM, leading to the intercalation site, and TS2, 2.4 kcal mol⁻¹ above IM, leading to the minor-groove site.

Figure 3 (top) illustrates representative structures along the metadynamics intercalated to minor-groove-bound pathway. These were obtained as the average structures of clusters⁶⁰ lying within 2 Å rmsd of each significant point (transition state or free energy minimum) along the pathway. Looking again at Figure 2, we see that the drug initially moves mainly in the X -direction, since its anthraquinone moiety is still held between the DNA base pairs. Once this interaction weakens, the drug begins to rotate (decreasing θ) and DNA bends away from

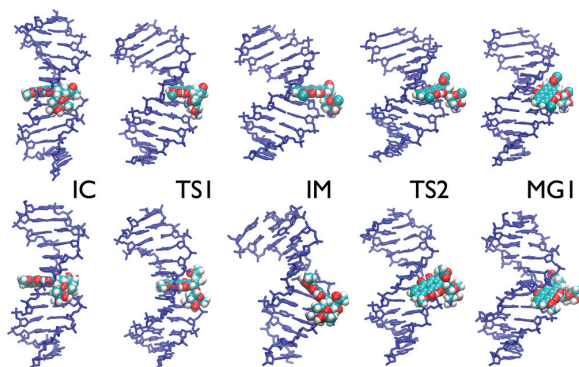


Figure 3. Representative structures leading from the intercalation site to the minor-groove-bound site for the metadynamics (above) and umbrella-sampling (below) pathways. DNA is shown as a blue Dreiding model and daunomycin as a CPK model with standard chemical coloring.

daunomycin, creating a large positive roll at the CpG step in order to regain some favorable stacking between these base pairs (see the top central image of Figure 3). This is the situation at the intermediate IM state. As the drug moves slightly further away and continues to rotate, DNA can relax by recovering full CpG stacking, while at $\theta \approx 80^\circ$, the drug finally locates its optimal position within the minor groove. The main conformational parameters of the states along the metadynamics pathway are given in Table 2A.

Table 2. Structural Comparison of Significant Points along the Free Energy Pathways from the Intercalated State (IC), through the Intermediate (IM) State, to the Minor-Groove-Bound State (MG1)^a

Parameters	IC	TS1	IM	TS2	MG1
(A) Metadynamics					
rise C6pG7 (Å)	7.4	7.3	6.0	5.5	3.8
roll C6pG7 (deg)	-2	28	44	42	8
Σ roll A5-T8 (deg)	-2	40	45	44	7
twist C6pG7 (deg)	38	25	28	29	39
Σ twist A5-T8 (deg)	81	74	84	84	99
minor-groove width (Å)	9.5	10.4	10.9	10.5	6.9
A-ring pucker	eq.	eq.	eq.	ax.	ax.
(B) Umbrella Sampling					
rise (deg)	7.4	7.2	4.2	3.6	3.8
roll C6pG7 (deg)	0	6	39	2	-4
Σ roll A5-T8 (deg)	-9	8	32	6	-4
twist C6pG7 (deg)	34	24	14	35	41
Σ twist A5-T8 (deg)	87	77	74	95	99
minor-groove width (Å)	7.0	9.0	11.3	6.0	4.9
A-ring pucker	eq.	strained	ax.	ax.	eq.

^aAll parameters were calculated using Curves+.⁵⁸ The minor-groove width is measured midway between the C6 and G7 base pairs.

Umbrella Sampling of Daunomycin Conformations.

Since the metadynamics pathway showed that the conformation of daunomycin changes as the drug moves from the intercalation site to the groove (see Table 2A), we began the umbrella sampling studies by analyzing the internal flexibility of daunomycin and, notably, the properties of its nonplanar A-ring (see Figure 1). It has been shown experimentally that this ring favors two pucker states that can be characterized by the torsion

angle τ (C7–C8–C9–C10). $\tau \approx -60^\circ$ places the C13 acetyl group in an axial position, while $\tau \approx +50^\circ$ places it in an equatorial position.^{61–65} The most recent experiments find that the equatorial state is favored in solution, although the axial state is observed in the presence of divalent ions.⁶⁶

We have studied the repuckering of the A-ring in three different environments: free in aqueous solution, in the intercalated state (IC), and in the minor-groove-bound state (MG1). The corresponding free energy curves were obtained by umbrella sampling of the torsion angle τ , initially using minimal salt conditions as for the metadynamics studies described in the Metadynamics Intercalation Pathway section. As expected, they all show an axial minimum close to $\tau = -60^\circ$ and an equatorial minimum close to $\tau = +50^\circ$. When daunomycin is isolated (Figure 4A), these two states are almost equally stable and are separated by a barrier of ≤ 3 kcal mol⁻¹. In contrast, the minor-groove-bound state (Figure 4B) shows a clear preference for the axial conformation (1.5 kcal mol⁻¹), coupled with a larger barrier separating the two states (≥ 4 kcal mol⁻¹). Lastly, in the intercalated site, the DNA environment strongly favors an equatorial conformation (Figure 4C), in agreement with crystallographic observations,³¹ and any values of $\tau < +20^\circ$ are disfavored by at least 6 kcal mol⁻¹.

During this study, we noticed that changing the A-ring puckering in the minor-groove-bound state led to a strong coupling with the minor-groove width, which increased by roughly 3 Å as daunomycin transitioned to the wider axial state (Supporting Information, Figure S2A). This led us to wonder whether minimal salt conditions were exaggerating the role of interactions between DNA and the charged daunosamine sugar of the drug. We consequently repeated the free energy calculations using a 0.15 M NaCl concentration, close to physiological conditions. This indeed led to a decoupling of the minor-groove width (see Figure S2B, Supporting Information) and significant change in the free energy profile (see Figure 4B, blue curve), which now favored the equatorial conformation of the A-ring. The free energy profiles for free or intercalated daunomycin with a physiological salt concentration showed less change and did not affect the preferred A-ring pucker. However, given the impact on the minor-groove width, we decided to use physiological salt conditions for the umbrella-sampling studies of the daunomycin–DNA intercalation pathway.

Umbrella Sampling Intercalation Pathway. Umbrella sampling was used to construct a single pathway between the intercalated state and the groove-bound state from two separate pathways, one removing daunomycin from its intercalation site (the IC pathway) and the other removing daunomycin from its minor-groove-bound site (the MG1 pathway). These two pathways were found to intersect and can thus be used to build a single IC–MG1 free energy profile.

The IC pathway removed daunomycin from its intercalation site using a separation restraint acting on the rigid part of the A-ring (atoms C7–C6a–C10a–C10 in Figure 1) and the atoms of the two base pairs forming the intercalation site. This choice ensured movement toward the minor-groove side of the intercalation site, while leaving the ligand free to reorient once it was no longer stacked with the flanking bases. This restraint had a value of 3.2 Å for the IC state. Umbrella sampling was carried out over the range 3.1–12.5 Å in 0.15 Å steps with 4–8 ns of sampling at each point (more sampling was used in

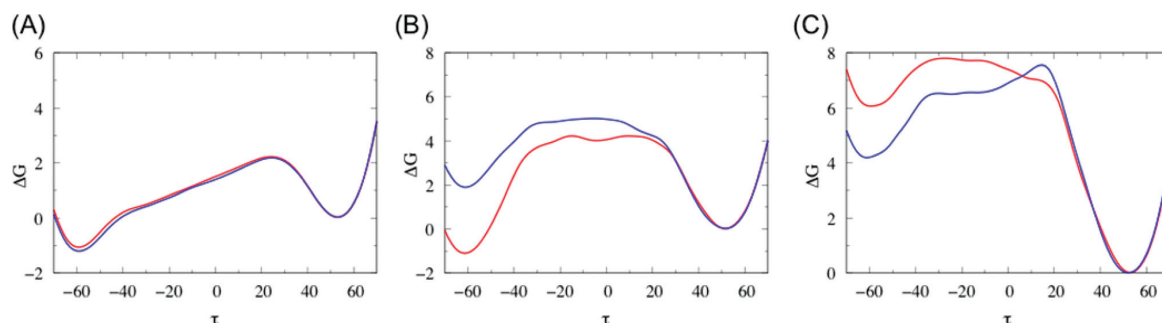


Figure 4. Free energy profiles (kcal mol^{-1}) for repuckering the A-ring of daunomycin using the dihedral $\tau = \text{C7–C8–C9–C10}$ (deg). Local free energy minima are found at $\tau \approx 50^\circ$ (corresponding to an equatorial position of the C13 acetyl group) and at $\tau \approx -60^\circ$ (corresponding to an axial position of the acetyl group): (A) daunomycin in aqueous solution, (B) daunomycin bound to the minor groove of DNA, and (C) daunomycin intercalated within at the C6pG7 step of DNA. In each case, the red curve was obtained in minimal salt conditions and the blue curve in “physiological” conditions (0.15 M NaCl).

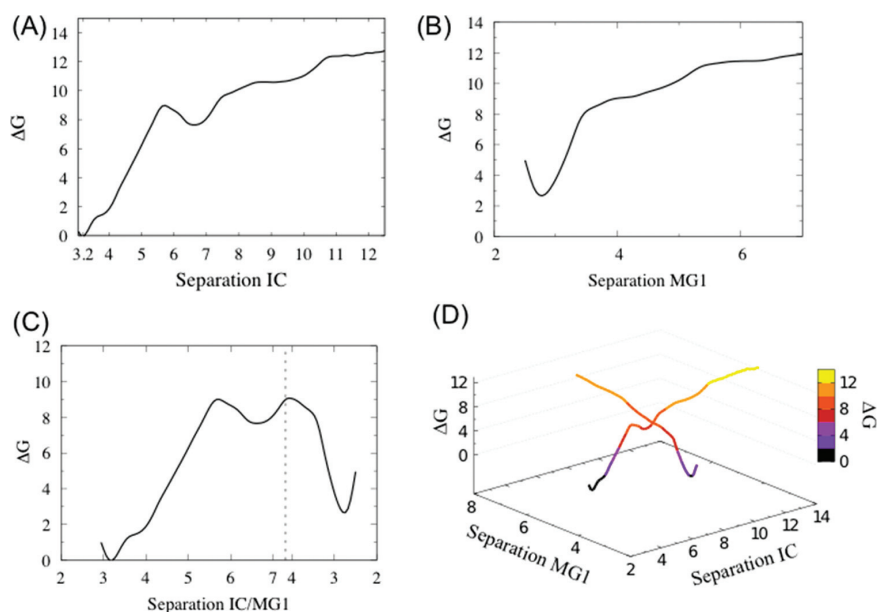


Figure 5. Umbrella-sampling free energy pathways (energies in kcal mol^{-1} and separation distances in Å): (A) separation from the intercalated state (IC), showing the intermediate state (IM); (B) separation from the minor-groove-bound state (MG1); (C) assembly of a single pathway between IC and MG1 at the point indicated by the vertical dotted line; and (D) 3D view of the intersection of the two free energy pathways.

regions where large conformational changes occurred either in DNA or in the orientation of the drug).

The resulting free energy profile (Figure 5A) shows a sharp, linear free energy rise as the anthraquinone ring system is removed from the DNA base stack. At 5.8 Å, daunomycin begins to rotate, reflected by a local free energy maximum 9 kcal mol^{-1} above the free energy of the intercalated state, followed by a metastable intermediate state at 6.8 Å, with a free energy of 7.7 kcal mol^{-1} . This intermediate lies within a quadratic free energy well that ends at roughly 7.4 Å. Further separation of the drug leads to a slower, more or less linear free energy rise. The final free energy is 13 kcal mol^{-1} above the IC state, but it is important to note that complete sampling is very difficult once the ligand moves away from DNA and is free to tumble.

A 2D rmsd plot of representative snapshots drawn from this free energy profile (Figure 6A) confirms the existence of five

conformational clusters along the pathway: (a) intercalated or partially intercalated, (b) first transition state, (c) metastable intermediate, (d) second transition state, and (e) separation. Note that the intermediate state flanked by two transition states was not observed in our earlier, less detailed study of the reaction pathway.³²

A second umbrella-sampling pathway, the MG1 pathway, was used to remove daunomycin from the minor-groove-bound state that we constructed during our earlier studies by manually placing the drug in the center of the minor groove and then running unrestrained molecular dynamics.³² In order to remove the drug from the groove we used a separation restraint acting on all heavy atoms of the ligand and on all base atoms of DNA. The restraint had a value of 2.8 Å for the MG1 state. Umbrella sampling was carried out over the range 2.5–7.1 Å in 0.15 Å steps with 6 ns of MD sampling at each step. Figure 5B shows an approximately quadratic free energy profile around the

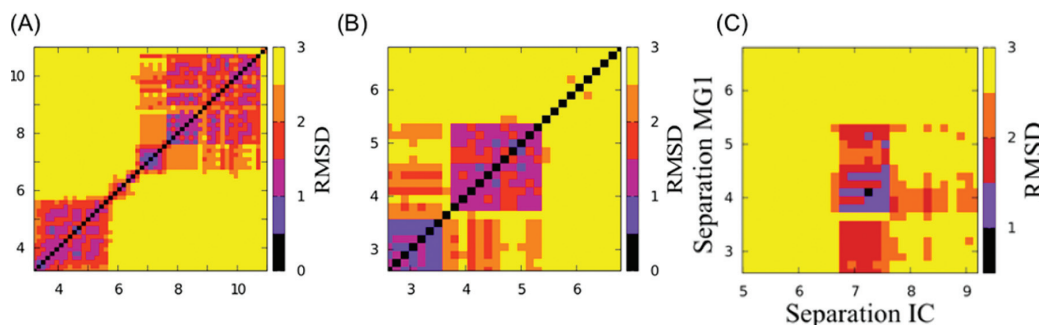


Figure 6. 2D rmsd plots (Å) between DNA–daunomycin conformations drawn from the umbrella-sampling pathways. The numbers along the axes show the increasing separation restraints, and smaller rmsd differences are shown by darker colors. Dark patches centered on the diagonal indicate similar structures that persist for a range of separation distances: (A) conformations along the pathway from the intercalated state (IC), (B) conformations along the pathway from the minor-groove-bound state (MG1), and (C) between conformations from the two pathways, showing the common structures sampled at 7.25 Å along the IC pathway and at 4.1 Å along the MG1 pathway.

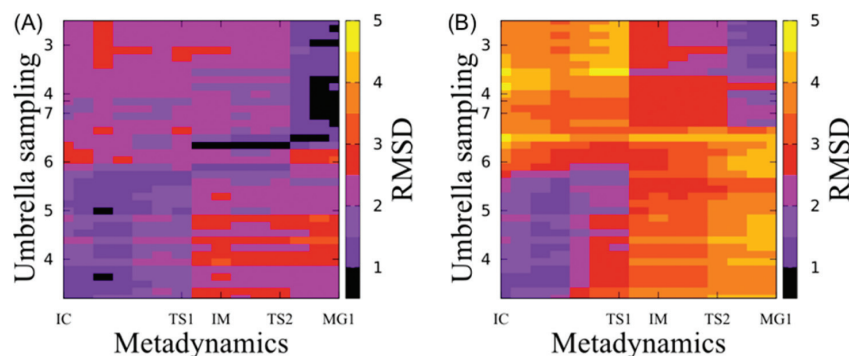


Figure 7. 2D rmsd plots (Å) between conformations drawn from the metadynamics pathway (horizontal axis) and the umbrella-sampling pathway (vertical axis): (A) for DNA alone and (B) for the DNA–daunomycin complex. See the caption to Figure 6 for further explanations.

minimum, followed by a more linear free energy rise starting from a separation of 3.5 Å, when attractive interactions between daunomycin and the sides of the minor groove are lost. The changes that occur at this point are reflected in the 2D rmsd plot shown in Figure 6B.

As shown by the 2D rmsd plot in Figure 6C, it is in fact possible to link the IC and MG1 pathways into a single IC–MG1 pathway. Indeed, we find a very close correspondence between the structures at a separation of 7.25 Å on the IC pathway and those at a separation of 4.1 Å on the MG1 pathway. We recall that these two distances differ significantly because they refer to different choices of interface atoms for the restraint.

Joining the two free energy pathways allows us to place the minor-groove minimum at a free energy 2.7 kcal mol^{−1} relative to the intercalated state (see Figures 5C,D). Representative structures along the IC–MG1 pathway, including the intermediate free energy minimum IM and the two flanking transition states, TS1 and TS2, are illustrated in Figure 3 (bottom line), and their main conformational parameters are given in Table 2B.

Comparison of Metadynamics and Umbrella Results.

A first comparison of the metadynamics and umbrella sampling IC–MG1 pathways can be made using the free energy values in Table 1 and the conformational parameters in Table 2. The free energies of the significant points along both pathways are very similar to one another. The differences are almost certainly below the accuracy of the calculations, with the possible

exception of the minor-groove-bound state MG1. The umbrella-sampling pathway finds this state to be 2.7 kcal mol^{−1} less stable than the intercalated state (compared to only 0.4 kcal mol^{−1} using metadynamics). We attribute this to the use of a physiological salt concentration for the umbrella-sampling simulations, since these conditions diminish the importance of the ionic interactions between the daunomycin sugar and DNA backbone. This result also seems more in line with the fact that the minor-groove-bound state is only an intermediate preceding intercalation. The physiological salt concentration may also be responsible for slightly lowering the free energies of the transition states on either side of the IM intermediate.

The conformational parameters for the states along the two pathways are given in Table 2, and as for the free energies discussed above, they follow similar trends. The IC and MG1 states are similar (except for a narrower minor groove for the umbrella-sampling MG1 state). Some differences occur for IM and the two flanking transition states, although these are not very evident from the conformational snapshots shown in Figure 3, apart from differences in the daunomycin orientation. For the IM state, the orientation of daunomycin varies significantly during the sampling along both pathways, notably in terms of the orientation of the drug (as shown in Figure S3, Supporting Information). Despite differences in the representative structures shown in Figure 3, similar structures are visited along both pathways (as shown in Figure S4, Supporting Information). Moreover, it should be noted that

despite movements of the drug, the DNA structures sampled are very similar with minimal and mean rmsd values of 0.3 and 1.4 Å for the nucleotides constituting the intercalation site. The differences in daunomycin orientation at IM may be linked to the change in the internal conformation of the A-ring (which is axial along the metadynamics pathway and equatorial along the umbrella sampling pathway; see Table 2) or may be linked to the different salt conditions used for the two pathways.

For a more global comparison of the metadynamics and umbrella-sampling results, we have performed two further analyses. The first uses 2D rmsd plots comparing representative conformations drawn from the two pathways. As Figure 7 shows, the agreement is good for the significant points along the pathway, both for the DNA alone (rmsd <2 Å) and for the DNA–daunomycin complex (rmsd <2.5 Å), with somewhat larger values for the IM state, where the inclination of the drug fluctuates significantly as discussed above.

The second, and more continuous, view of the pathways is given in the 3D plot in Figure 8 that combines information on

significant points along the pathway show that the metadynamics path (open circles) and the umbrella-sampling path (filled circles) are very similar in the chosen variables. The differences in the IM state due to daunomycin orientation mentioned above do not show up here, as only DNA–drug distance is plotted. These results also show that the transition states, TS1 and TS2, occur closer to IM in the metadynamics pathway than in the umbrella-sampling pathway. However, conformational changes occur very rapidly during these transitions, as witnessed by the sparsity of sampling between the IC, IM, and MG1 regions.

DISCUSSION AND CONCLUSIONS

By using two different free energy sampling methods we have obtained a consensus view of how the drug daunomycin intercalates within the base stack of a DNA double helix. The optimal pathway involves initial binding to the minor groove with minimal DNA deformation (with a binding free energy of roughly -10 kcal mol $^{-1}$ using the umbrella-sampling simulations with a physiological salt concentration). An activated process (requiring roughly 6.5 kcal mol $^{-1}$ under the same conditions) causes the drug to rotate and to partially insert its planar anthraquinone moiety into a wedge formed between two successive base pairs. This intermediate state lies in a shallow free energy well (with a depth of roughly 1.5–2 kcal mol $^{-1}$). Crossing a small free energy barrier is therefore sufficient to achieve full intercalation.

The existence of an intermediate state, where the drug is partially intercalated and DNA is wedged open on the minor-groove side and bent away from the drug, is of particular interest when we consider other DNA interactions. Similar states are seen in with number of minor-groove-binding proteins including high-mobility group proteins, such as SRY⁶⁷ and HMGB1,⁶⁸ with Sac7d,⁶⁹ and also with the TATA-box binding protein.⁷⁰ In each of these cases, the protein partially intercalates a hydrophobic side chain between successive base pairs, causing a positive roll angle and significant bending toward the major groove (see Figure S5, Supporting Information). A similar state occurs in the crystal structure of a partially intercalating polypyridyl ruthenium complex with DNA⁷¹ and, again, in very recent simulation studies of the interaction of Δ -[Ru(phenanthroline)₃]²⁺ with DNA in aqueous solution. The interesting contrast between all these cases and the IM state is that the overall conformation of the proteins, or of the ligands, prevents full intercalation. In contrast, for daunomycin, the IM state is only a shallow local minimum on the intercalation pathway. However, in the light of this discussion, it is perhaps not so surprising that the IM state exists.

From a therapeutic point of view, it is interesting to note that when DNA kinking toward the major groove is stabilized, for example, by cis-Pt binding to successive bases in the major groove, this can enhance protein binding in the minor groove, notably by HMGB1, which can then block DNA repair and induce cell apoptosis.^{68,72} This suggests that if it were possible to significantly stabilize the IM state by chemically modifying daunomycin, or related drugs, this might provide an interesting new direction for drug development.

Lastly, we consider the kinetic implications of the intercalation pathway. Following the direction in which the simulations were performed, we consider the daunomycin dissociation reaction. The simulations indicate that crossing a 9 kcal mol $^{-1}$ barrier is necessary to reach the IM intermediate

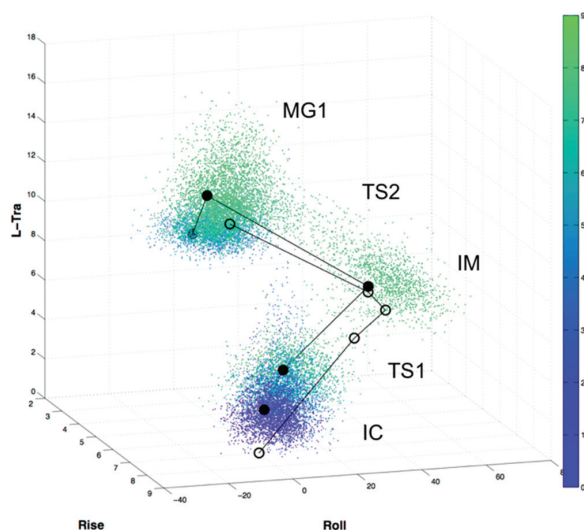


Figure 8. Pathway sampling from the minor-groove site to the intercalation site. The 3D plot shows the rise and roll of DNA at the C6pG7 intercalation step (horizontal plane) and the drug to center of DNA axis distance, L-Tra (vertical axis). Snapshots from the umbrella-sampling pathway are shown as points colored from blue to green (increasing free energy with respect to the intercalated state, see color bar in kcal mol $^{-1}$). The main states along the pathway are labeled IC (high rise and low roll), IM (high positive roll), and MG1 (low rise and high L-Tra), and TS1 (separating IC and IM) and TS2 (separating IM and MG1) are the less densely sampled arms joining the main clusters. Representative structures from the metadynamics and umbrella-sampling pathways are shown as open and filled circles, respectively.

DNA structure (namely, the rise and roll at the C6pG7 intercalation step), information on the ligand position (the distance from the center of the DNA helical axis), and free energy (colored blue to green as free energy increases). This scatter plot contains more than 15 000 snapshots from the umbrella-sampling IC–MG1 path which clearly visualize the clusters of points associated with the intercalated state IC (bottom), the intermediate state IM (center right), and the groove-bound state MG1 (top). The lines between the

state. A second small barrier of 1.5–2 kcal mol⁻¹ separates this state from the minor-groove-bound site. Roughly 10 kcal mol⁻¹ is then required to escape into solution. This result is partially compatible with the three-state model of Chaires.^{16,73} If we associate the “outside-bound” state (Chaires state 1) with our minor-groove-binding state, and the final state (Chaires state 3) with the intercalated state, then the activation energies (estimated from the temperature dependence of the corresponding rate constants) for escaping from the intercalation site (14.3 kcal mol⁻¹) and then from the minor-groove site (10.7 kcal mol⁻¹) fit with our data reasonably well. The second barrier to escape our intermediate state, which does not appear in the 3-state model,¹⁶ might contribute to the larger single activation energy seen for this step. However, we see no process related to the unimolecular step between Chaires states 2 and 3 corresponding to “a conformational rearrangement of the intercalated drug molecule or the DNA molecule” (with a high activation energy for escaping from this third state of 16.2 kcal mol⁻¹).¹⁶ It is possible that this step corresponds to pre-equilibrium repositioning of the drug at another intercalation site between different base pairs.³² In this respect, it should be stressed that, just as in our previous work,³² the present PMF of daunomycin/DNA interactions excludes any nonequilibrium features, since the aim is to equilibrate the system at each step along the reaction pathway. This could result in suppressing the rearrangement responsible for the final step in the Chaires model and could also hide intermediate steps in the five-state model of Rizzo et al. (who notably break down the Chaires step 2 → 3 into two constituent steps). Such pathway features could be investigated using nonequilibrium simulations in the future, in which case the present work would be a valuable starting point. Further studies will also be necessary to characterize drug movements along the minor groove and to analyze the role of water and ion dynamics around the daunomycin–DNA complex.

■ ASSOCIATED CONTENT

Supporting Information

Description of metadynamics collective variables, simulation protocols and error analysis, conformational analysis of ligand–DNA complexes, daunomycin conformational fluctuations, conformational details of the metastable partially intercalated intermediate state. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

arnab.mukherjee@iiserpune.ac.in; chynes43@gmail.com; r.lavery@ibcp.fr

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge funding from the ANR Blanc project ALADDIN and CINES for the allocation of supercomputer resources. B.B., K.Z., and R.L. thank the CNRS for their support. This work was supported in part by NSF grants CHE-0750477 and CHE-1112564 (J.T.H.). A.M. and J.T.H. acknowledge TeraGrid for supercomputing facilities. A.M. also thanks the computational facility of IISER, Pune, India.

■ REFERENCES

- (1) Tewey, K. M.; Rowe, T. C.; Yang, L.; Halligan, B. D.; Liu, L. F. *Science* **1984**, 226, 466–468.
- (2) Crow, R. T.; Crothers, D. M. *J. Med. Chem.* **1994**, 37, 3191–3194.
- (3) Denny, W. A. *Expert Opin. Invest. Drugs* **1997**, 6, 1845–1851.
- (4) Denny, W. A.; Baguley, B. C. *Curr. Top. Med. Chem.* **2003**, 3, 339–353.
- (5) Lerman, L. S. *J. Mol. Biol.* **1961**, 3, 18–30.
- (6) Strekowski, L.; Wilson, B. *Mutat. Res.* **2007**, 623, 3–13.
- (7) Weiss, R. B. *Semin. Oncol.* **1992**, 19, 670–686.
- (8) Minotti, G.; Menna, P.; Salvatorelli, E.; Cairo, G.; Gianni, L. *Pharmacol. Rev.* **2004**, 56, 185–229.
- (9) Müller, W.; Crothers, D. M. *J. Mol. Biol.* **1968**, 35, 251–290.
- (10) Chaires, J. B.; Dattagupta, N.; Crothers, D. M. *Biochemistry* **1982**, 21, 3933–3940.
- (11) Li, H. J.; Crothers, D. M. *J. Mol. Biol.* **1969**, 39, 461–477.
- (12) Chaires, J. B. *Biochemistry* **1983**, 22, 4204–4211.
- (13) Graves, D. E.; Krugh, T. R. *Biochemistry* **1983**, 22, 3941–3947.
- (14) Forster, W.; Stutter, E. *Int. J. Biol. Macromol.* **1984**, 6, 114–124.
- (15) Wilson, W. D.; Krishnamoorthy, C. R.; Wang, Y. H.; Smith, J. C. *Biopolymers* **1985**, 24, 1941–1961.
- (16) Chaires, J. B.; Dattagupta, N.; Crothers, D. M. *Biochemistry* **1985**, 24, 260–267.
- (17) Krishnamoorthy, C. R.; Yen, S. F.; Smith, J. C.; Lown, J. W.; Wilson, W. D. *Biochemistry* **1986**, 25, 5933–5940.
- (18) Macgregor, R. B. J.; Clegg, R. M.; Jovin, T. M. *Biochemistry* **1987**, 26, 4008–4016.
- (19) Rizzo, V.; Sacchi, N.; Menozzi, M. *Biochemistry* **1989**, 28, 274–282.
- (20) Ramstein, J.; Dourlent, M.; Leng, M. *Biochem. Biophys. Res. Commun.* **1972**, 47, 874–882.
- (21) Bresloff, J. L.; Crothers, D. M. *J. Mol. Biol.* **1975**, 95, 103–110.
- (22) Corin, A. F.; Jovin, T. M. *Biochemistry* **1986**, 25, 3995–4007.
- (23) Leger, J. F.; Robert, J.; Bourdieu, L.; Chatenay, D.; Marko, J. F. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, 95, 12295–12299.
- (24) DiMarco, A.; Gaetani, M.; Orezzi, P.; Scarpinato, B. M.; Silvestrini, R.; Soldati, M.; Dasdia, T.; Valentini, L. *Nature* **1964**, 201, 706–707.
- (25) Myers Jr, C. E.; Chabner, B. A. *Cancer Chemotherapy: Principles and Practice*; Lippincott: Philadelphia, PA, 1990.
- (26) Tsai, C. C.; Jain, S. C.; Sobell, H. M. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, 72, 628–632.
- (27) Quigley, G. J.; Wang, A. H.; Ughetto, G.; Van Der Marel, G.; Van Boom, J. H.; Rich, A. *Proc. Natl. Acad. Sci. U. S. A.* **1980**, 77, 7204–7208.
- (28) Wang, A. H. J.; Ughetto, G.; Quigley, G. J.; Rich, A. *Biochemistry* **1987**, 26, 1152–1163.
- (29) Nunn, C. M.; Van Meervelt, L.; Zhang, S.; Moore, M. H.; Kennard, O. *J. Mol. Biol.* **1991**, 222, 167–177.
- (30) Davies, D. B.; Eaton, R. J.; Baranovsky, S. F.; Veselkov, A. N. *J. Biomol. Struct. Dyn.* **2000**, 17, 887–901.
- (31) Barthwal, R.; Sharma, U.; Srivastava, N.; Jain, M.; Awasthi, P.; Kaur, M.; Barthwal, S. K.; Govil, G. *Eur. J. Med. Chem.* **2006**, 41, 27–39.
- (32) Mukherjee, A.; Lavery, R.; Bagchi, B.; Hynes, J. T. *J. Am. Chem. Soc.* **2008**, 130, 9747–9755.
- (33) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99, 12562–12566.
- (34) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, 127, 2600–2607.
- (35) Pietrucci, F.; Marinelli, F.; Carloni, P.; Laio, A. *J. Am. Chem. Soc.* **2009**, 131, 11811–11818.
- (36) Vargiu, A. V.; Ruggerone, P.; Magistrato, A.; Carloni, P. *Nucleic Acids Res.* **2008**, 36, 5910–5921.
- (37) Bouvier, B.; Lavery, R. *J. Am. Chem. Soc.* **2009**, 131, 9864–9865.
- (38) Bouvier, B.; Zakrzewska, K.; Lavery, R. *Angew. Chem., Int. Ed.* **2011**, 50, 6516–6518.

- (39) Bewley, C. A.; Gronenborn, A. M.; Clore, G. M. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 105–131.
- (40) Lavery, R.; Zakrzewska, K.; Sklenar, H. *Comput. Phys. Commun.* **1995**, *91*, 135–158.
- (41) Chen, K. X.; Gresh, N.; Pullman, B. *J. Biomol. Struct. Dyn.* **1985**, *3*, 445–466.
- (42) Chaires, J. B.; Herrera, J. E.; Waring, M. J. *Biochemistry* **1990**, *29*, 6145–6153.
- (43) Roche, C. J.; Thomson, J. A.; Crothers, D. M. *Biochemistry* **1994**, *33*, 926–935.
- (44) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (45) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.
- (46) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T.; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J. *Nucleic Acids Res.* **2010**, *38*, 299–313.
- (47) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (48) Besler, B. H.; Merz, K. M., Jr; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431–439.
- (49) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C. *Gaussian*; Gaussian, Inc.: Wallingford, CT, 2003.
- (50) Jorgensen, W. L.; Jenson, C. J. *Comput. Chem.* **1998**, *19*, 1179–1186.
- (51) Dang, L. X. *J. Am. Chem. Soc.* **1995**, *117*, 6954–6960.
- (52) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (53) Bhatele, A.; Kumar, S.; Mei, C.; Phillips, J. C.; Zheng, G.; Kale, L. V. 2009, .
- (54) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302–238306.
- (55) Camilloni, C.; Provati, D.; Tiana, G.; Broglia, R. A. *Proteins: Struct., Funct. Bioinf.* **2008**, *71*, 1647–1654.
- (56) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (57) McLachlan, A. D. *Eur. J. Biochem.* **1979**, *100*, 181–187.
- (58) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. *Nucleic Acids Res.* **2009**, *37*, S917–S929.
- (59) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38.
- (60) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; Van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236–240.
- (61) Nuss, M. E.; James, T. L.; Apple, M. A.; Kollman, P. A. *Biochim. Biophys. Acta-Nucleic Acids Protein Synth.* **1980**, *609*, 136–147.
- (62) Lown, J. W.; Chen, H. H. *Can. J. Chem.* **1981**, *59*, 3212–3217.
- (63) Mondelli, R.; Ragg, E.; Fronza, G.; Arnone, A. *J. Chem. Soc., Perkin Trans. 2* **1987**, 15–26.
- (64) Barthwal, R.; Mujeeb, A.; Srivastava, N.; Sharma, U. *Chem.–Biol. Interact.* **1996**, *100*, 125–139.
- (65) Barthwal, R.; Srivastava, N.; Sharma, U.; Govil, G. *J. Mol. Struct.* **1994**, *327*, 201–220.
- (66) Myung, J. M.; Jhon, M. S.; Kang, Y. K. *Bull. Korean Chem. Soc.* **1987**, *8*, 39–45.
- (67) King, C. Y.; Weiss, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 11990–11994.
- (68) Ohndorf, U. M.; Rould, M. A.; He, Q.; Pabo, C. O.; Lippard, S. J. *Nature* **1999**, *399*, 708–712.
- (69) Robinson, H.; Gao, Y. G.; McCrary, B. S.; Edmondson, S. P.; Shriver, J. W.; Wang, A. H. *Nature* **1998**, *392*, 202–205.
- (70) Kim, Y.; Geiger, J. H.; Hahn, S.; Sigler, P. B. *Nature* **1993**, *365*, 512–520.
- (71) Hall, J. P.; O’Sullivan, K.; Naseer, A.; Smith, J. A.; Kelly, J. M.; Cardin, C. J. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17610–17614.
- (72) Wang, D.; Zhu, G.; Huang, X.; Lippard, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 9584–9589.
- (73) Chaires, J. B. *Biophys. Chem.* **1990**, *35*, 191–202.

ANNEXE C

Recouvrement des échantillonnages

Figure C.1 – Recouvrement des échantillonnages le long du chemin de sortie du site d'intercalation. Chaque gaussienne représente les variations des valeurs de contraintes au sein d'une fenêtre de simulation.

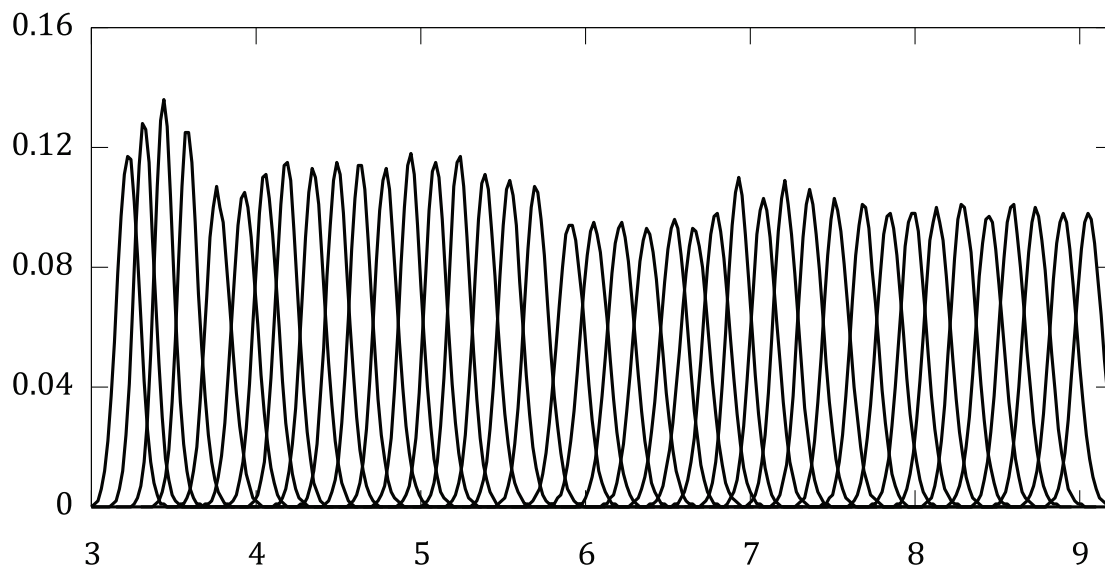
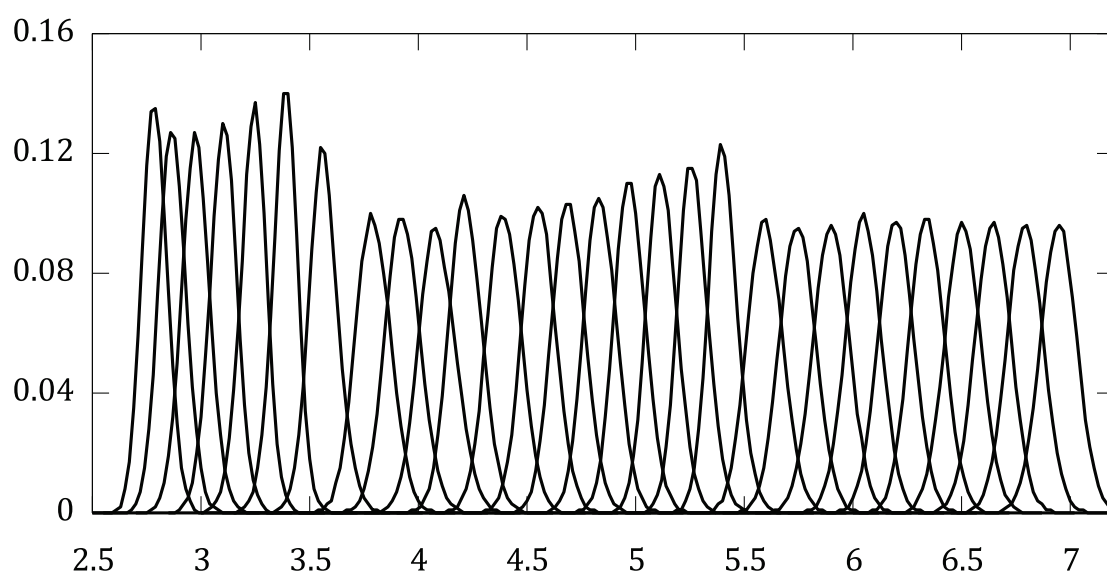


Figure A. 2 – Recouvrement des échantillonnages le long du chemin de séparation de la DNM liée au petit sillon.



Bibliographie

1. Miescher, F. Ueber die chemische Zusammensetzung der Eiterzellen. *Medicinisch-chemische Untersuchungen* **4**, 441-460 (1871).
2. Kossel, A. Ueber die basischen Stoffe des Zellkerns. *Hoppe-Seyler's Zeitschrift für physiologische Chemie* **22**, 176-187 (1897).
3. Jones, W. Nucleic acids: their chemical properties and physiological conduct. *Longmans, Green, London* (1920).
4. Feulgren, R. & Rossenbeck, H. Mikroskopisch-chemischer Nachweis einer Nucleinsäure vom Typus der Thymonucleinsäure und die-darauf beruhende elective Färbung von Zellkernen in mikroskopischen Präparaten. *Hoppe-Seyler's Zeitschrift für physiologische Chemie* **135**, 203-248 (1924).
5. Avery, O. T., MacLeod, C. M. & McCarty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *J Exp Med* 137-158 (1944).
6. Chargaff, E. & Vischer, E. The composition of the desoxypentose nucleic acids of thymus and spleen. *J Biol Chem* **177**, 405-416 (1949).
7. Chargaff, E. Structure and function of nucleic acids as cell constituents. *Fed Proc* **10**, 654-659 (1951).
8. Hershey, A. D. & Chase, M. Independent functions of viral protein and nucleic acid in growth of bacteriophageage. *The Journal of General Physiology* **36**, 39-56 (1952).
9. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids. *Nature* **171**, 737-738 (1953).

10. Altona, C. & Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *Journal of the American Chemical Society* **94**, 8205-8212 (1972).
11. Murray-Rust, P. & Motherwell, S. Computer retrieval and analysis of molecular geometry. III. Geometry of the β -1'-aminofuranoside fragment. *Acta Crystallographica Section B* **34**, 2534-2546 (1978).
12. Kennard, O. & Hunter, W. N. Landolt-Bornstein Numerical Data and Functional Relationships in Science and Technology, Group VII/1a, Nucleic Acids (1989).
13. Djuranovic, D. & Hartmann, B. DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers* **73**, 356-368 (2004).
14. Dickerson, R. E. & Drew, H. R. Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *J Mol Biol* **149**, 761-786 (1981).
15. Dickerson, R. E. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res* **17**, 1797-1803 (1989).
16. Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., *et al.* A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol* **313**, 229-237 (2001).
17. Baraldi, P. G., Bovero, A., Fruttarolo, F., Preti, D., *et al.* DNA minor groove binders as potential antitumor and antimicrobial agents. *Med Res Rev* **24**, 475-528 (2004).
18. Pabo, C. O. & Sauer, R. T. Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem* **61**, 1053-1095 (1992).
19. Hong, M. & Marmorstein, R. in *Protein-nucleic acid interactions: structural biology* 47-65 (Royal Society of Chemistry, 2008).
20. Boutonnet, N., Hui, X. & Zakrzewska, K. Looking into the grooves of DNA. *Biopolymers* **33**, 479-490 (1993).
21. Seeman, N. C., Rosenberg, J. M. & Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proceedings of the National Academy of Sciences* **73**, 804 (1976).

22. Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., *et al.* Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* **335**, 321-329 (1988).
23. Haran, T. E. & Mohanty, U. The unique structure of A-tracts and intrinsic DNA bending. *Q Rev Biophys* **42**, 41-81 (2009).
24. Hunt, R. A., Munde, M., Kumar, A., Ismail, M. A., *et al.* Induced topological changes in DNA complexes: influence of DNA sequences and small molecule structures. *Nucleic Acids Res* **39**, 4265-4274 (2011).
25. Lankas, F., Spackova, N., Moakher, M., Enkhbayar, P. & Sponer, J. A measure of bending in nucleic acids structures applied to A-tract DNA. *Nucleic Acids Research* **38**, 3414-3422 (2010).
26. Nelson, H. C., Finch, J. T., Luisi, B. F. & Klug, A. The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* **330**, 221-226 (1987).
27. Stefl, R., Wu, H., Ravindranathan, S., Sklenár, V. & Feigon, J. DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4 conundrum. *Proc Natl Acad Sci U S A* **101**, 1177-1182 (2004).
28. Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M. & Zhurkin, V. B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc Natl Acad Sci U S A* **95**, 11163-11168 (1998).
29. Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, R. M., *et al.* The anatomy of A-, B-, and Z-DNA. *Science* **216**, 475-485 (1982).
30. Arnott, S. & Hukins, D. W. Optimised parameters for A-DNA and B-DNA. *Biochem Biophys Res Commun* **47**, 1504-1509 (1972).
31. Arnott, S. & Hukins, D. W. Refinement of the structure of B-DNA and implications for the analysis of x-ray diffraction data from fibers of biopolymers. *J Mol Biol* **81**, 93-105 (1973).
32. Arnott, S. & Selsing, E. Letter: The structure of polydeoxyguanylic acid with polydeoxycytidylic acid. *J Mol Biol* **88**, 551-552 (1974).
33. Pilet, J. & Brahms, J. Dependence of B-A conformational change in DNA on base composition. *Nat New Biol* **236**, 99-100 (1972).

34. Chandrasekaran, R. & Arnott, S. in Landolt-Bornstein Numerical Data and Functional Relationships in Science and Technology, Group VII/1b, Nucleic Acids (Springer, Berlin, 1989).
35. Schneider, B., Neidle, S. & Berman, H. M. Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers* **42**, 113-124 (1997).
36. Leslie, A. G., Arnott, S., Chandrasekaran, R. & Ratliff, R. L. Polymorphism of DNA double helices. *J Mol Biol* **143**, 49-72 (1980).
37. Kennard, O., Cruse, W. B., Nachman, J., Prange, T., *et al.* Ordered water structure in an A-DNA octamer at 1.7 Å resolution. *J Biomol Struct Dyn* **3**, 623-647 (1986).
38. Shakked, Z., Guenstein-Guzikevich, G., Eisenstein, M., Frolow, F. & Rabinovich, D. The conformation of the DNA double helix in the crystal is dependent on its environment. *Nature* **342**, 456-460 (1989).
39. Wang, A. H., Quigley, G. J., Kolpak, F. J., Crawford, J. L., *et al.* Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680-686 (1979).
40. Herbert, A. & Rich, A. Left-handed Z-DNA: structure and function. *Genetica* **106**, 37-47 (1999).
41. Arnott, S., Bond, P. J. & Chandrasekaran, R. Visualization of an unwound DNA duplex. *Nature* **287**, 561-563 (1980).
42. Ha, S. C., Lowenhaupt, K., Rich, A., Kim, Y. G. & Kim, K. K. Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* **437**, 1183-1186 (2005).
43. Zhao, J., Bacolla, A., Wang, G. & Vasquez, K. M. Non-B DNA structure-induced genetic instability and evolution. *Cell Mol Life Sci* **67**, 43-62 (2010).
44. Dickerson, R. E. & Klug, A. Base sequence and helix structure variation in B and A DNA. *Journal of Molecular Biology* **166**, 419 - 441 (1983).
45. Zhurkin, V. B., Lysov, Y. P. & Ivanov, V. I. Anisotropic flexibility of DNA and the nucleosomal structure. *Nucleic Acids Res* **6**, 1081-1096 (1979).

46. Dickerson, R. E. & Chiu, T. K. Helix bending as a factor in protein/DNA recognition. *Biopolymers* **44**, 361-403 (1997).
47. Koudelka, G. B., Mauro, S. A. & Ciubotaru, M. Indirect readout of DNA sequence by proteins: the roles of DNA sequence-dependent intrinsic and extrinsic forces. *Prog Nucleic Acid Res Mol Biol* **81**, 143-177 (2006).
48. Locasale, J. W., Napoli, A. A., Chen, S., Berman, H. M. & Lawson, C. L. Signatures of protein-DNA recognition in free DNA binding sites. *Journal of molecular biology* **386**, 1054-1065 (2009).
49. Calladine, C. R. Mechanics of sequence-dependent stacking of bases in B-DNA. *J Mol Biol* **161**, 343-352 (1982).
50. el Hassan, M. A. & Calladine, C. R. Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J Mol Biol* **259**, 95-103 (1996).
51. Yanagi, K., Privé, G. G. & Dickerson, R. E. Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J Mol Biol* **217**, 201-214 (1991).
52. Bax, A., Kontaxis, G. & Tjandra, N. Dipolar couplings in macromolecular structure determination. *Methods Enzymol* **339**, 127-174 (2001).
53. MacDonald, D. & Lu, P. Residual dipolar couplings in nucleic acid structure determination. *Current opinion in structural biology* **12**, 337-343 (2002).
54. Park, H. J. DNA structural perturbation induced by the CPI-derived DNA interstrand cross-linker: molecular mechanisms for the sequence specific recognition. *Arch Pharm Res* **24**, 455-465 (2001).
55. Wu, Z., Delaglio, F., Tjandra, N., Zhurkin, V. B. & Bax, A. Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and ³¹P chemical shift anisotropy. *J Biomol NMR* **26**, 297-315 (2003).
56. Koo, H. S., Wu, H. M. & Crothers, D. M. DNA bending at adenine extperiod-centered thymine tracts. *Nature* **320**, 501-506 (1986).
57. Crothers, D. M. & Shakked, Z. DNA bending by adenine-thymine tracts. *Oxford Handbook of Nucleic Acid Structure. Oxford University Press, Oxford* 455-470 (1998).

58. Hizver, J., Rozenberg, H., Frolow, F., Rabinovich, D. & Shakked, Z. DNA bending by an adenine--thymine tract and its role in gene regulation. *Proc Natl Acad Sci U S A* **98**, 8490-8495 (2001).
59. Zhurkin, V. B., Tolstorukov, M. Y., Xu, F., Colasanti, A. V. & Olson, W. K. Sequence-Dependent Variability of B-DNA. *DNA conformation and transcription* 18-34 (2005).
60. Goodsell, D. S., Kaczor-Grzeskowiak, M. & Dickerson, R. E. The crystal structure of C-C-A-T-T-A-A-T-G-G. Implications for bending of B-DNA at T-A steps. *J Mol Biol* **239**, 79-96 (1994).
61. Ulyanov, N. B. & Zhurkin, V. B. Sequence-dependent anisotropic flexibility of B-DNA. A conformational study. *J Biomol Struct Dyn* **2**, 361-385 (1984).
62. Olson, W. K. & Zhurkin, V. B. Modeling DNA deformations. *Curr Opin Struct Biol* **10**, 286-297 (2000).
63. Gorin, A. A., Zhurkin, V. B. & Olson, W. K. B-DNA twisting correlates with base-pair morphology. *J Mol Biol* **247**, 34-48 (1995).
64. Bertrand, H., Ha-Duong, T., Fermandjian, S. & Hartmann, B. Flexibility of the B-DNA backbone: effects of local and neighbouring sequences on pyrimidine-purine steps. *Nucleic Acids Res* **26**, 1261-1267 (1998).
65. Bandyopadhyay, D. & Bhattacharyya, D. Effect of neighboring bases on base-pair stacking orientation: a molecular dynamics study. *J Biomol Struct Dyn* **18**, 29-43 (2000).
66. Brukner, I., Sánchez, R., Suck, D. & Pongor, S. Trinucleotide models for DNA bending propensity: comparison of models based on DNaseI digestion and nucleosome packaging data. *J Biomol Struct Dyn* **13**, 309-317 (1995).
67. Rohs, R., West, S. M., Liu, P. & Honig, B. Nuance in the double-helix and its role in protein-DNA recognition. *Curr Opin Struct Biol* **19**, 171-177 (2009).
68. Orozco, M., Noy, A. & Pérez, A. Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr Opin Struct Biol* **18**, 185-193 (2008).
69. Fujii, S., Kono, H., Takenaka, S., Go, N. & Sarai, A. Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic Acids Res* **35**, 6063-6074 (2007).

70. Gardiner, E. J., Hunter, C. A., Packer, M. J., Palmer, D. S. & Willett, P. Sequence-dependent DNA structure: a database of octamer structural parameters. *J Mol Biol* **332**, 1025-1035 (2003).
71. Gardiner, E. J., Hunter, C. A., Lu, X. J. & Willett, P. A structural similarity analysis of double-helical DNA. *J Mol Biol* **343**, 879-889 (2004).
72. Beveridge, D. L., Barreiro, G., Byun, K. S., Case, D. A., *et al.* Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys J* **87**, 3799-3813 (2004).
73. Dixit, S. B., Beveridge, D. L., Case, D. A., Cheatham, T. E., *et al.* Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys J* **89**, 3721-3740 (2005).
74. Lavery, R., Zakrzewska, K., Beveridge, D., Bishop, T. C., *et al.* A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res* **38**, 299-313 (2010).
75. Pérez, A., Marchán, I., Svozil, D., Sponer, J., *et al.* Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J* **92**, 3817-3829 (2007).
76. Cheatham, T. E., Cieplak, P. & Kollman, P. A. A modified version of the Cornell *et al.* force field with improved sugar pucker phases and helical repeat. *J Biomol Struct Dyn* **16**, 845-862 (1999).
77. Auffinger, P. & Westhof, E. Water and ion binding around RNA and DNA (C,G) oligomers. *J Mol Biol* **300**, 1113-1131 (2000).
78. Shakked, Z. The influence of the environment on DNA structures determined by X-ray crystallography. *Current Opinion in Structural Biology* **1**, 446-451 (1991).
79. Shen, X., Gu, B., Che, S. A. & Zhang, F. S. Solvent effects on the conformation of DNA dodecamer segment: a simulation study. *J Chem Phys* **135**, 34509-34520 (2011).

Bibliographie

80. Auffinger, P. & Westhof, E. Roles of hydration on the structure and dynamics of nucleic acids. *Water management in the design and distribution of quality foods* 165-198 (1999).
81. Westhof, E. Water: an integral part of nucleic acid structure. *Annual review of biophysics and biophysical chemistry* **17**, 125-144 (1988).
82. Westhof, E. & Beveridge, D. L. Hydration of nucleic acids. *The Molecules of life* **5**, 24 (1991).
83. Auffinger, P. & Hashem, Y. Nucleic acid solvation: from outside to insight. *Curr Opin Struct Biol* **17**, 325-333 (2007).
84. Ball, P. Water as an active constituent in cell biology. *Chem Rev* **108**, 74-108 (2008).
85. Chalikian, T. V. & Breslauer, K. J. Thermodynamic analysis of biomolecules: a volumetric approach. *Curr Opin Struct Biol* **8**, 657-664 (1998).
86. Schneider, B. & Berman, H. Basics of Nucleic Acid Structure. *Computational Studies of RNA and DNA* 1-44 (2006).
87. Franklin, R. E. & Gosling, R. G. The structure of sodium thymonucleate fibres. I. The influence of water content. *Acta Crystallographica* **6**, 673-677 (1953).
88. Schneider, B., Cohen, D. & Berman, H. M. Hydration of DNA bases: analysis of crystallographic data. *Biopolymers* **32**, 725-750 (1992).
89. Schneider, B., Cohen, D. M., Schleifer, L., Srinivasan, A. R., *et al.* A systematic method for studying the spatial distribution of water molecules around nucleic acid bases. *Biophys J* **65**, 2291-2303 (1993).
90. Schneider, B. & Berman, H. M. Hydration of the DNA bases is local. *Biophysical journal* **69**, 2661-2669 (1995).
91. Schneider, B., Patel, K. & Berman, H. M. Hydration of the phosphate group in double-helical DNA. *Biophys J* **75**, 2422-2434 (1998).
92. Drew, H. R. & Dickerson, R. E. Structure of a B-DNA dodecamer. III. Geometry of hydration. *J Mol Biol* **151**, 535-556 (1981).
93. Kopka, M. L., Fratini, A. V., Drew, H. R. & Dickerson, R. E. Ordered water structure around a B-DNA dodecamer. A quantitative study. *J Mol Biol* **163**, 129-146 (1983).

94. Mack, D. R., Chiu, T. K. & Dickerson, R. E. Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: a comparative analysis of T-A and A-T steps within A-tracts. *J Mol Biol* **312**, 1037-1049 (2001).
95. Shatzky-Schwartz, M., Arbuckle, N. D., Eisenstein, M., Rabinovich, D., *et al.* X-ray and solution studies of DNA oligomers and implications for the structural basis of A-tract-dependent curvature. *J Mol Biol* **267**, 595-623 (1997).
96. Tereshko, V., Minasov, G. & Egli, M. A “hydrat-ion” spine in a B-DNA minor groove. *Journal of the American Chemical Society* **121**, 3590-3595 (1999).
97. Rohs, R., Jin, X., West, S. M., Joshi, R., *et al.* Origins of specificity in protein-DNA recognition. *Annu Rev Biochem* **79**, 233-269 (2010).
98. Chuprina, V. P., Heinemann, U., Nurislamov, A. A., Zielenkiewicz, P., *et al.* Molecular dynamics simulation of the hydration shell of a B-DNA decamer reveals two main types of minor-groove hydration depending on groove width. *Proc Natl Acad Sci U S A* **88**, 593-597 (1991).
99. Wahl, M. C., Rao, S. T. & Sundaralingam, M. Crystal structure of the B-DNA hexamer d(CTCGAG): model for an A-to-B transition. *Biophys J* **70**, 2857-2866 (1996).
100. Arai, S., Chatake, T., Ohhara, T., Kurihara, K., *et al.* Complicated water orientations in the minor groove of the B-DNA decamer d(CCATTAATGG)₂ observed by neutron diffraction measurements. *Nucleic Acids Res* **33**, 3017-3024 (2005).
101. Yonetani, Y., Maruyama, Y., Hirata, F. & Kono, H. Comparison of DNA hydration patterns obtained using two distinct computational methods, molecular dynamics simulation and three-dimensional reference interaction site model theory. *J Chem Phys* **128**, 185102 (2008).
102. Cheatham III, T. E. & Kollman, P. A. Molecular dynamics simulations highlight the structural differences among DNA: DNA, RNA: RNA, and DNA: RNA hybrid duplexes. *Journal of the American Chemical Society* **119**, 4805-4825 (1997).
103. Young, M. A., Ravishanker, G. & Beveridge, D. L. A 5-nanosecond molecular dynamics trajectory for B-DNA: analysis of structure, motions, and solvation. *Biophys J* **73**, 2313-2336 (1997).

104. Reddy, S. Y., Leclerc, F. & Karplus, M. DNA polymorphism: a comparison of force fields for nucleic acids. *Biophys J* **84**, 1421-1449 (2003).
105. Toporowski, J. W., Reddy, S. Y. & Bruice, T. C. An investigation of the ionic and solvation patterns of dsDNG versus dsDNA by use of molecular dynamics simulations. *Biophys Chem* **126**, 132-139 (2007).
106. Feig, M. & Pettitt, B. M. A molecular simulation picture of DNA hydration around A- and B-DNA. *Biopolymers* **48**, 199-209 (1998).
107. Manning, G. S. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys* **11**, 179-246 (1978).
108. Hud, N. V. & Feigon, J. Localization of divalent metal ions in the minor groove of DNA A-tracts. *Journal of the American Chemical Society* **119**, 5756-5757 (1997).
109. Hud, N. V., Sklenár, V. & Feigon, J. Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending. *J Mol Biol* **286**, 651-660 (1999).
110. Chad, C., McFail-Isom, L., Howerton, S. B., VanDerveer, D. & Williams, L. D. Cations Mediate B-DNA Conformational Heterogeneity. *Journal of the American Chemical Society* **122**, 11048-11056 (2000).
111. Howard, J. J., Lynch, G. C. & Pettitt, B. M. Ion and solvent density distributions around canonical B-DNA from integral equations. *J Phys Chem B* **115**, 547-556 (2011).
112. Howerton, S. B., Sines, C. C., VanDerveer, D. & Williams, L. D. Locating monovalent cations in the grooves of B-DNA. *Biochemistry* **40**, 10023-10031 (2001).
113. Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry* **37**, 8341-8355 (1998).
114. Shui, X., Sines, C. C., McFail-Isom, L., VanDerveer, D. & Williams, L. D. Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry* **37**, 16877-16887 (1998).

115. Hud, N. V. & Polak, M. DNA-cation interactions: The major and minor grooves are flexible ionophores. *Curr Opin Struct Biol* **11**, 293-301 (2001).
116. Sletten, E. & Frøystein, N. A. NMR studies of oligonucleotide-metal ion interactions. *Met Ions Biol Syst* **32**, 397-418 (1996).
117. Cheatham, T. E. & Kollman, P. A. Molecular dynamics simulation of nucleic acids. *Annu Rev Phys Chem* **51**, 435-471 (2000).
118. Nishimura, Y., Torigoe, C. & Tsuboi, M. Salt induced B—A transition of poly (dG). poly (dC) and the stabilization of A form by its methylation. *Nucleic acids research* **14**, 2737-2748 (1986).
119. Feig, M. & Pettitt, B. M. Sodium and chlorine ions as part of the DNA solvation shell. *Biophys J* **77**, 1769-1781 (1999).
120. Heddi, B., Foloppe, N., Hantz, E. & Hartmann, B. The DNA structure responds differently to physiological concentrations of K(+) or Na(+). *J Mol Biol* **368**, 1403-1411 (2007).
121. Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *Journal of Biological chemistry* **257**, 14686-14707 (1982).
122. Grzeskowiak, K., Yanagi, K., Privé, G. G. & Dickerson, R. E. The structure of B-helical C-G-A-T-C-G-A-T-C-G and comparison with C-C-A-A-C-G-T-T-G-G. The effect of base pair reversals. *J Biol Chem* **266**, 8861-8883 (1991).
123. Privé, G. G., Yanagi, K. & Dickerson, R. E. Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. *J Mol Biol* **217**, 177-199 (1991).
124. el Antri, S., Bittoun, P., Mauffret, O., Monnot, M., *et al.* Effect of distortions in the phosphate backbone conformation of six related octanucleotide duplexes on CD and ³¹P NMR spectra. *Biochemistry* **32**, 7079-7088 (1993).
125. el Antri, S., Mauffret, O., Monnot, M., Lescot, E., *et al.* Structural deviations at CpG provide a plausible explanation for the high frequency of mutation at this site. Phosphorus nuclear magnetic resonance and circular dichroism studies. *J Mol Biol* **230**, 373-378 (1993).

126. Gorenstein, D. G. Conformation and Dynamics of DNA and Protein-DNA Complexes by ³¹P NMR. *Chemical Reviews* **94**, 1315-1338 (1994).
127. Djuranovic, D. & Hartmann, B. Molecular dynamics studies on free and bound targets of the bovine papillomavirus type I e2 protein: the protein binding effect on DNA and the recognition mechanism. *Biophys J* **89**, 2542-2551 (2005).
128. Hartmann, B., Sullivan, M. R. & Harris, L. F. Operator recognition by the phage 434 cI repressor: MD simulations of free and bound 50-bp DNA reveal important differences between the OR1 and OR2 sites. *Biopolymers* **68**, 250-264 (2003).
129. Tisné, C., Delepierre, M. & Hartmann, B. How NF-kappaB can be attracted by its cognate DNA. *J Mol Biol* **293**, 139-150 (1999).
130. Wellenzohn, B., Flader, W., Winger, R. H., Hallbrucker, A., *et al.* Indirect readout of the trp-repressor-operator complex by B-DNA's backbone conformation transitions. *Biochemistry* **41**, 4088-4095 (2002).
131. Djuranovic, D. & Hartmann, B. Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: evidence for sub-states. *J Biomol Struct Dyn* **20**, 771-788 (2003).
132. Hartmann, B., Piazzola, D. & Lavery, R. BI-BII transitions in B-DNA. *Nucleic acids research* **21**, 561-568 (1993).
133. Heddi, B., Foloppe, N., Bouchemal, N., Hantz, E. & Hartmann, B. Quantification of DNA BI/BII backbone states in solution. Implications for DNA overall structure and recognition. *J Am Chem Soc* **128**, 9170-9177 (2006).
134. Varnai, P., Djuranovic, A., Lavery, R. & Hartmann, B. a/g Transitions in the B-DNA backbone. *Nucleic Acids Research* **30**, 5398-5406 (2002).
135. Ravishanker, G., Swaminathan, S., Beveridge, D. L., Lavery, R. & Sklenar, H. Conformational and helicoidal analysis of 30 PS of molecular dynamics on the d(CGCGAATTCGCG) double helix: "curves", dials and windows. *J Biomol Struct Dyn* **6**, 669-699 (1989).
136. Schmitz, U., Ulyanov, N. B., Kumar, A. & James, T. L. Molecular Dynamics with Weighted Time-averaged Restraints for a DNA Octamer : Dynamic Interpreta-

tion of Nuclear Magnetic Resonance Data. *Journal of Molecular Biology* **234**, 373-389 (1993).

137. Várnai, P. & Zakrzewska, K. DNA and its counterions: a molecular dynamics study. *Nucleic Acids Res* **32**, 4269-4280 (2004).

138. Weisz, K., Shafer, R. H., Egan, W. & James, T. L. Solution structure of the octamer motif in immunoglobulin genes via restrained molecular dynamics calculations. *Biochemistry* **33**, 354-366 (1994).

139. Boer, D. R., Canals, A. & Coll, M. DNA-binding drugs caught in action: the latest 3D pictures of drug-DNA complexes. *Dalton Trans* 399-414 (2009).

140. Hurley, L. H. DNA and its associated processes as targets for cancer therapy. *Nat Rev Cancer* **2**, 188-200 (2002).

141. Fichou, Y. & Férec, C. The potential of oligonucleotides for therapeutic applications. *Trends Biotechnol* **24**, 563-570 (2006).

142. François, J. C., Saison-Behmoaras, T. & Hélène, C. Sequence-specific recognition of the major groove of DNA by oligodeoxynucleotides via triple helix formation. Footprinting studies. *Nucleic Acids Res* **16**, 11431-11440 (1988).

143. Wartell, R. M., Larson, J. E. & Wells, R. D. Netropsin. *Journal of Biological Chemistry* **249**, 6719-6731 (1974).

144. Lippert, B. Cisplatin, Chemistry and Biochemistry of a Leading Anti-Cancer Drug. *Wiley-VCH*. (1999).

145. Komeda, S., Moulaei, T., Woods, K. K., Chikuma, M., *et al.* A third mode of DNA binding: Phosphate clamps by a polynuclear platinum complex. *J Am Chem Soc* **128**, 16092-16103 (2006).

146. Lerman, L. S. Structural considerations in the interaction of DNA and acridines. *J Mol Biol* **3**, 18-30 (1961).

147. Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. & Ratliff, R. L. Left-handed DNA helices. *Nature* **283**, 743-745 (1980).

148. Bond, P. J., Langridge, R., Jennette, K. W. & Lippard, S. J. X-ray fiber diffraction evidence for neighbor exclusion binding of a platinum metallointercalation reagent to DNA. *Proc Natl Acad Sci U S A* **72**, 4825-4829 (1975).

149. Chothers, D. M. Calculation of binding isotherms for heterogeneous polymers. *Biopolymers* **6**, 575-584 (1968).
150. Gueron, M. & Leroy, J. L. The i-motif in nucleic acids. *Current opinion in structural biology* **10**, 326-331 (2000).
151. Gehring, K., Leroy, J. L. & Guéron, M. A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* **363**, 561-565 (1993).
152. Laugaa, P., Markovits, J., Delbarre, A., Le Pecq, J. B. & Roques, B. P. DNA tris intercalation: first acridine trimer with DNA affinity in the range of DNA regulatory proteins. Kinetic studies. *Biochemistry* **24**, 5567-5575 (1985).
153. Lee, J., Guelev, V., Sorey, S., Hoffman, D. W. & Iverson, B. L. NMR structural analysis of a modular threading tetraintercalator bound to DNA. *J Am Chem Soc* **126**, 14036-14042 (2004).
154. Murr, M. M., Harting, M. T., Guelev, V., Ren, J., *et al.* An octakis-intercalating molecule. *Bioorganic & Medicinal Chemistry* **9**, 1141-1148 (2001).
155. Goodman, L. S. & Wintrobe, M. M. Nitrogen mustard therapy; use of methyl-bis (beta-chloroethyl) amine hydrochloride and tris (beta-chloroethyl) amine hydrochloride for Hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders. *J Am Med Assoc* **132**, 126-132 (1946).
156. Lerman, L. S. Acridine mutagens and DNA structure. *Journal of Cellular and Comparative Physiology* **64**, 1-18 (1964).
157. Martinez, R. & Chacon-Garcia, L. The search of DNA-intercalators as anti-tumoral drugs: what it worked and what did not work. *Current medicinal chemistry* **12**, 127-151 (2005).
158. Hande, K. R. Clinical applications of anticancer drugs targeted to topoisomerase II. *Biochim Biophys Acta* **1400**, 173-184 (1998).
159. Hande, K. R. Topoisomerase II inhibitors. *Update on Cancer Therapeutics* **3**, 13-26 (2008).
160. Neto, B. A. & Lapis, A. A. Recent developments in the chemistry of deoxyribonucleic acid (DNA) intercalators: principles, design, synthesis, applications and trends. *Molecules* **14**, 1725-1746 (2009).

161. Dimarco, A., Gaetani, M., Orezzi, P., Scarpinato, B. M., *et al.* 'Daunomycin', a new antibiotic of the rhodomycin group. *Nature* **201**, 706-707 (1964).
162. Dubost, M., Ganter, P., Maral, R., Ninet, L., *et al.* Rubidomycin: a new antibiotic with cytostatic properties. *Cancer Chemother Rep* **41**, 35-36 (1964).
163. Barone, G., Guerra, C. F., Gambino, N., Silvestri, A., *et al.* Intercalation of daunomycin into stacked DNA base pairs. DFT study of an anticancer drug. *J Biomol Struct Dyn* **26**, 115-130 (2008).
164. Gewirtz, D. A. A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. *Biochem Pharmacol* **57**, 727-741 (1999).
165. Minotti, G., Menna, P., Salvatorelli, E., Cairo, G. & Gianni, L. Anthracyclines: molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity. *Pharmacol Rev* **56**, 185-229 (2004).
166. Crow, R. T. & Crothers, D. M. Inhibition of topoisomerase I by anthracycline antibiotics: evidence for general inhibition of topoisomerase I by DNA-binding agents. *J Med Chem* **37**, 3191-3194 (1994).
167. Chaires, J. B. in *Small Molecule DNA and RNA Binders* 461-481 (Wiley-VCH Verlag GmbH & Co. KGaA, 2004).
168. Frederick, C. A., Williams, L. D., Ughetto, G., van der Marel, G. A., *et al.* Structural comparison of anticancer drug-DNA complexes: adriamycin and daunomycin. *Biochemistry* **29**, 2538-2549 (1990).
169. Nunn, C. M., Van Meervelt, L., Zhang, S. D., Moore, M. H. & Kennard, O. DNA-drug interactions. The crystal structures of d(TGTACA) and d(TGATCA) complexed with daunomycin. *J Mol Biol* **222**, 167-177 (1991).
170. Davies, D. B., Eaton, R. J., Baranovsky, S. F. & Veselkov, A. N. NMR investigation of the complexation of daunomycin with deoxytetranucleotides of different base sequence in aqueous solution. *J Biomol Struct Dyn* **17**, 887-901 (2000).
171. Barthwal, R., Sharma, U., Srivastava, N., Jain, M., *et al.* Structure of daunomycin complexed to d-TGATCA by two-dimensional nuclear magnetic resonance spectroscopy. *Eur J Med Chem* **41**, 27-39 (2006).

172. Chaires, J. B., Dattagupta, N. & Crothers, D. M. Studies on interaction of anthracycline antibiotics and deoxyribonucleic acid: equilibrium binding studies on interaction of daunomycin with deoxyribonucleic acid. *Biochemistry* **21**, 3933-3940 (1982).
173. Suh, D., Oh, Y. K., Hur, M. W., Ahn, B. & Chaires, J. B. Daunomycin binding to deoxypolynucleotides with alternating sequences: complete thermodynamic profiles of heterogeneous binding sites. *Nucleosides Nucleotides Nucleic Acids* **21**, 637-649 (2002).
174. Xodo, L. E., Manzini, G., Ruggiero, J. & Quadrifoglio, F. On the interaction of daunomycin with synthetic alternating DNAs: sequence specificity and polyelectrolyte effects on the intercalation equilibrium. *Biopolymers* **27**, 1839-1857 (1988).
175. Pachter, J. A., Huang, C. H., DuVernay, V. H., Prestayko, A. W. & Crooke, S. T. Viscometric and fluorometric studies of deoxyribonucleic acid interactions of several new anthracyclines. *Biochemistry* **21**, 1541-1547 (1982).
176. Dalglish, D. G., Fey, G. & Kersten, W. Circular dichroism studies of complexes of the antibiotics daunomycin, nogalamycin, chromomycin, and mithramycin with DNA. *Biopolymers* **13**, 1757-1766 (1974).
177. Chaires, J. B., Fox, K. R., Herrera, J. E., Britt, M. & Waring, M. J. Site and sequence specificity of the daunomycin-DNA interaction. *Biochemistry* **26**, 8227-8236 (1987).
178. Courseille, C. ., Busetta, B. ., Geoffre, S. . & Hospital, M. . Complex daunomycin-butanol. *Acta Crystallographica Section B* **35**, 764-767 (1979).
179. Neidle, S. & Taylor, G. Nucleic acid binding drugs. Part IV. The crystal structure of the anti-cancer agent daunomycin. *Biochim Biophys Acta* **479**, 450-459 (1977).
180. Neidle, S. & Taylor, G. L. Nucleic acid binding drugs. Some conformational properties of the anti-cancer drug daunomycin and several of its derivatives: implications for DNA-binding. *FEBS Lett* **107**, 348-354 (1979).
181. Wang, A. H. J., Ughetto, G., Quigley, G. J. & Rich, A. Interactions between an anthracycline antibiotic and DNA: molecular structure of daunomycin com-

plexed to d (CpGpTpApCpG) at 1.2-Å resolution. *Biochemistry* **26**, 1152-1163 (1987).

182. Clark, G. R., Pytel, P. D., Squire, C. J. & Neidle, S. Structure of the first parallel DNA quadruplex-drug complex. *J Am Chem Soc* **125**, 4066-4067 (2003).

183. Gao, Y. -G., Robinson, H., Wijsman, E. R., van der Marel, G. A., *et al.* Binding of Daunorubicin to $\hat{\text{I}}^2$ -d-Glucosylated DNA Found in Protozoa *Trypanosoma brucei* Studied by X-ray Crystallography. *Journal of the American Chemical Society* **119**, 1496-1497 (1997).

184. Leonard, G. A., Hambley, T. W., McAuley-Hecht, K., Brown, T. & Hunter, W. N. Anthracycline-DNA interactions at unfavourable base-pair triplet-binding sites: structures of d(CGGCCG)/daunomycin and d(TGGCCA)/adriamycin complexes. *Acta Crystallogr D Biol Crystallogr* **49**, 458-467 (1993).

185. Moon, M. J. Effects of Hydration and Metal Ions on the Conformation of Daunomycin. *Bull. Korean Chem Soc.* **8**, 39-45 (1987).

186. Moore, M. H., Hunter, W. N., d'Estaintot, B. L. & Kennard, O. DNA-drug interactions. The crystal structure of d(CGATCG) complexed with daunomycin. *J Mol Biol* **206**, 693-705 (1989).

187. Schuerman, G. S. & Van Meervelt, L. Conformational Flexibility of the DNA Backbone. *Journal of the American Chemical Society* **122**, 232-240 (2000).

188. Shi, K., Pan, B. & Sundaralingam, M. Structure of a B-form DNA/RNA chimera (dC)(rG)d(ATCG) complexed with daunomycin at 1.5 Å resolution. *Acta Crystallogr D Biol Crystallogr* **59**, 1377-1383 (2003).

189. Fritzsche, H., Triebel, H., Chaires, J. B., Dattagupta, N. & Crothers, D. M. Studies on interaction of anthracycline antibiotics and deoxyribonucleic acid: geometry of intercalation of iremycin and daunomycin. *Biochemistry* **21**, 3940-3946 (1982).

190. Tsai, C. C., Jain, S. C. & Sobell, H. M. X-ray crystallographic visualization of drug-nucleic acid intercalative binding: structure of an ethidium-dinucleoside monophosphate crystalline complex, Ethidium: 5-iodouridylyl (3'-5') adenosine. *Proc Natl Acad Sci U S A* **72**, 628-632 (1975).

191. Waring, M. Variation of the supercoils in closed circular DNA by binding of antibiotics and drugs: Evidence for molecular models involving intercalation. *Journal of Molecular Biology* **54**, 247-279 (1970).
192. Trieb, M., Rauch, C., Wibowo, F. R., Wellenzohn, B. & Liedl, K. R. Cooperative effects on the formation of intercalation sites. *Nucleic Acids Res* **32**, 4696-4703 (2004).
193. Brown, S. C., Kollman, P. A. & Weiner, P. K. A conformational analysis of daunomycin considering D-ring repucker. *Biochim Biophys Acta* **717**, 49-55 (1982).
194. Nuss, M. E., James, T. L., Apple, M. A. & Kollman, P. A. An NMR study of the interaction of daunomycin with dinucleotides and dinucleoside phosphates. *Biochim Biophys Acta* **609**, 136-147 (1980).
195. Mondelli, R. & Ragg, E. Conformational Analysis of N-Acetyl-daunomycin in Solution. A Transient ¹H Nuclear Overhauser Effect Study of the Glycosidic Linkage Geometry. *J. Chem. Soc. Perkin Trans II* **1**, 27-32 (1987).
196. Mondelli, R. Nuclear Magnetic Resonance Conformational Study of Daunomycin and Related Antitumour Antibiotics in Solution. The Conformation of Ring. *J. Chem. Soc. Perkin Trans II* (1987).
197. Barthwal, R., Mujeeb, A. & Govil, G. Interaction of daunomycin with deoxydinucleotide d-CpG by two-dimensional proton magnetic resonance techniques. *Arch Biochem Biophys* **313**, 189-205 (1994).
198. Barthwal, R., Srivastava, N., Sharma, U. & Govil, G. A 500 MHz proton NMR study of the conformation of adriamycin. *Journal of Molecular Structure* **327**, 201-203 (1994).
199. Barthwal, R., Mujeeb, A., Srivastava, N. & Sharma, U. A proton nuclear magnetic resonance investigation of the conformation of daunomycin. *Chem Biol Interact* **100**, 125-139 (1996).
200. Barthwal, R., Agrawal, P., Tripathi, A. N., Sharma, U., *et al.* Structural elucidation of 4'-epiadriamycin by nuclear magnetic resonance spectroscopy and comparison with adriamycin and daunomycin using quantum mechanical and restrained molecular dynamics approach. *Arch Biochem Biophys* **474**, 48-64 (2008).

201. Schwarz. Rules for Conformation Nomenclature for Five- and Six-membered Rings in Monosaccharides and their Derivatives. *J.C.S. Chem. Comm* **14**, 505-508 (1973).
202. Lown, J. W. & Chen, H. H. Electron paramagnetic resonance characterization and conformation of daunorubicin semiquinone intermediate implicated in anthracycline metabolism, cardiotoxicity, and anticancer action. *Can. J. Chem.* **59**, 3212-3217 (1981).
203. Ren, J., Jenkins, T. C. & Chaires, J. B. Energetics of DNA intercalation reactions. *Biochemistry* **39**, 8439-8447 (2000).
204. Spolar, R. S. & Record, M. T. Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777-784 (1994).
205. Berman, H. M. & Young, P. R. The interaction of intercalating drugs with nucleic acids. *Annu Rev Biophys Bioeng* **10**, 87-114 (1981).
206. Waring, M. J. DNA modification and cancer. *Annual review of biochemistry* **50**, 159-192 (1981).
207. Chaires, J. B. A thermodynamic signature for drug-DNA binding mode. *Arch Biochem Biophys* **453**, 26-31 (2006).
208. Amzel, L. M. Loss of translational entropy in binding, folding, and catalysis. *Proteins: Structure, Function, and Bioinformatics* **28**, 144-149 (1997).
209. Amzel, L. M. Calculation of entropy changes in biological processes: folding, binding, and oligomerization. *Methods Enzymol* **323**, 167-177 (2000).
210. Chaires, J. B. Energetics of drug-DNA interactions. *Biopolymers* **44**, 201-215 (1997).
211. Haq, I., Jenkins, T. C., Chowdhry, B. Z., Ren, J. & Chaires, J. B. Parsing free energies of drug-DNA interactions. *Methods Enzymol* **323**, 373-405 (2000).
212. Record, M. T., Anderson, C. F. & Lohman, T. M. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Q Rev Biophys* **11**, 103-178 (1978).
213. Chaires, J. B. Dissecting the free energy of drug binding to DNA. *Anticancer Drug Des* **11**, 569-580 (1996).

Bibliographie

214. Jen-Jacobson, L., Engler, L. E. & Jacobson, L. A. Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure* **8**, 1015-1023 (2000).
215. Baginski, M., Fogolari, F. & Briggs, J. M. Electrostatic and non-electrostatic contributions to the binding free energies of anthracycline antibiotics to DNA. *J Mol Biol* **274**, 253-267 (1997).
216. Shui, X., Peek, M. E., Lipscomb, L. A., Wilkinson, A. P., *et al.* Effects of cationic charge on three-dimensional structures of intercalative complexes: structure of a bis-intercalated DNA complex solved by MAD phasing. *Curr Med Chem* **7**, 59-71 (2000).
217. Chaires, J. B., Satyanarayana, S., Suh, D., Fokt, I., *et al.* Parsing the free energy of anthracycline antibiotic binding to DNA. *Biochemistry* **35**, 2047-2053 (1996).
218. Remeta, D. P., Mudd, C. P., Berger, R. L. & Breslauer, K. J. Thermodynamic characterization of daunomycin-DNA interactions: Comparison of complete binding profiles for a series of DNA host duplexes. *Biochemistry* **32**, 5064-5073 (1993).
219. Chaires, J. B., Dattagupta, N. & Crothers, D. M. Kinetics of the daunomycin--DNA interaction. *Biochemistry* **24**, 260-267 (1985).
220. Rizzo, V., Sacchi, N. & Menozzi, M. Kinetic studies of anthracycline-DNA interaction by fluorescence stopped flow confirm a complex association mechanism. *Biochemistry* **28**, 274-282 (1989).
221. Trieb, M., Rauch, C., Wellenzohn, B., Wibowo, F., *et al.* Daunomycin intercalation stabilizes distinct backbone conformations of DNA. *J Biomol Struct Dyn* **21**, 713-724 (2004).
222. Honig, B. & Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **268**, 1144-1149 (1995).
223. Jayaram, B., Sprous, D. & Beveridge, D. L. Solvation free energy of biomacromolecules: Parameters for a modified generalized Born model consistent with the AMBER force field. *The Journal of Physical Chemistry B* **102**, 9571-9576 (1998).

224. Srinivasan, J., Cheatham III, T. E., Cieplak, P., Kollman, P. A. & David, A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *Journal of the American Chemical Society* **120**, 9401-9409 (1998).
225. Swanson, J. M., Henchman, R. H. & McCammon, J. A. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophysical Journal* **86**, 67-74 (2004).
226. Zou, X., Sun, Y. & Kuntz, I. D. Inclusion of solvation in ligand binding free energy calculations using the generalized-born model. *Journal of the American Chemical Society* **121**, 8033-8043 (1999).
227. Mukherjee, A., Lavery, R., Bagchi, B. & Hynes, J. T. On the molecular mechanism of drug intercalation into DNA: a simulation study of the intercalation pathway, free energy, and DNA structural changes. *J Am Chem Soc* **130**, 9747-9755 (2008).
228. McCammon, J. A. Protein dynamics. *Reports On Progress In Physics* **47**, 1-46 (1984).
229. Leach, A. R. *Molecular modelling: principles and applications* (Addison-Wesley Longman Ltd, 2001).
230. Kollman, P. Molecular modeling. *Annual Review of Physical Chemistry* **38**, 303-316 (1987).
231. van Gunsteren, W. F. & Berendsen, H. J. C. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angewandte Chemie International Edition in English* **29**, 992-1023 (1990).
232. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc* **117**, 5179-5197 (1995).
233. Cornell, W. D., Cieplak, P., Bayly, C. I. & Kollmann, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *Journal of the American Chemical Society* **115**, 9620-9631 (1993).
234. McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585-590 (1977).

235. Levitt, M. Computer simulation of DNA double-helix dynamics. *Cold Spring Harb. Symp. Quant. Biol* **47**, 251-262 (1983).
236. Tidor, B., Irikura, K. K., Brooks, B. R. & Karplus, M. Dynamics of DNA oligomers. *J Biomol Struct Dyn* **1**, 231-252 (1983).
237. Cheatham, T. E. I. I., Miller, J. L., Fox, T., Darden, T. A. & Kollman, P. A. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *Journal of the American Chemical Society* **117**, 4193-4194 (1995).
238. Darden, T. A., York, D. M. & Pedersen, L. G. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. *The Journal of chemical physics* **99**, 8345-8348 (1993).
239. Langley, D. R. Molecular dynamic simulations of environment and sequence dependent DNA conformations: the development of the BMS nucleic acid force field and comparison with experimental results. *J Biomol Struct Dyn* **16**, 487-509 (1998).
240. MacKerell, A. D., Wiorkiewicz-Kuczera, J. & Karplus, M. An all-atom empirical energy function for the simulation of nucleic acids. *Journal of the American Chemical Society* **117**, 11946-11975 (1995).
241. Foloppe, N. & MacKerell, J. ., A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of Computational Chemistry* **21**, 86-104 (2000).
242. Beveridge, D. L. & McConnell, K. J. Nucleic acids: theory and computer simulation, Y2K. *Curr Opin Struct Biol* **10**, 182-196 (2000).
243. Giudice, E. & Lavery, R. Simulations of nucleic acids and their complexes. *Acc Chem Res* **35**, 350-357 (2002).
244. Pérez, A., Blas, J. R., Rueda, M., López-Bes, J. M., *et al.* Exploring the Essential Dynamics of B-DNA. *Journal of Chemical Theory and Computation* **1**, 790-800 (2005).

245. Arthanari, H., McConnell, K. J., Beger, R., Young, M. A., *et al.* Assessment of the molecular dynamics structure of DNA in solution based on calculated and observed NMR NOESY volumes and dihedral angles from scalar coupling constants. *Biopolymers* **68**, 3-15 (2003).
246. Baker, N. A. Improving implicit solvent simulations: a Poisson-centric view. *Current Opinion in Structural Biology* **15**, 137-143 (2005).
247. Cramer, C. J. & Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chemical Reviews* **99**, 2161-2200 (1999).
248. Feig, M. & Brooks, C. L. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Current Opinion in Structural Biology* **14**, 217-224 (2004).
249. Simonson, T. Macromolecular electrostatics: continuum models and their growing pains. *Current Opinion in Structural Biology* **11**, 243-252 (2001).
250. Lazaridis, T. & Karplus, M. Thermodynamics of protein folding: a microscopic view. *Biophysical Chemistry* **100**, 367-395 (2003).
251. Sosa, C. P., Hewitt, T., Lee, M. R. & Case, D. A. Vectorization of the generalized Born model for molecular dynamics on shared-memory computers. *Journal of Molecular Structure: THEOCHEM* **549**, 193-201 (2001).
252. Gabdouliline, R. R. & Wade, R. C. Brownian Dynamics Simulation of Protein-Protein Diffusional Encounter. *Methods* **14**, 329-341 (1998).
253. Madura, J. D., Briggs, J. M., Wade, R. C., Davis, M. E., *et al.* Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Computer Physics Communications* **91**, 57-95 (1995).
254. Huber, G. A. & McCammon, J. A. Browndye: A Software Package for Brownian Dynamics. *Comput Phys Commun* **181**, 1896-1905 (2010).
255. Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., *et al.* Anton, a special-purpose machine for molecular dynamics simulation. *ACM SIGARCH Computer Architecture News* **35**, 1-12 (2007).
256. Chocholousová, J. & Feig, M. Implicit solvent simulations of DNA and DNA-protein complexes: agreement with explicit solvent vs experiment. *J Phys Chem B* **110**, 17240-17251 (2006).

257. Tsui, V. & David, A. Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *Journal of the American Chemical Society* **122**, 2489-2498 (2000).
258. Gaillard, T. & Case, D. A. Evaluation of DNA Force Fields in Implicit Solvation. *Journal of Chemical Theory and Computation* **7**, 3181-3198 (2011).
259. Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *Journal of the American Chemical Society* **103**, 335-340 (1981).
260. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926-935 (1983).
261. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F. & Hermans, J. Interaction models for water in relation to protein hydration. *Intermolecular forces* 331-342, (1981).
262. Stillinger, F. H. & Rahman, A. Improved simulation of liquid water by molecular dynamics. *The Journal of Chemical Physics* **60**, 1545-1557 (1974).
263. Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics* **112**, 8910-8922 (2000).
264. Wang, J. & Cieplak, P. kollman PA: How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules. *J Comput Chem* **21**, 1049-1074 (2000).
265. Aqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *The Journal of Physical Chemistry* **94**, 8021-8024 (1990).
266. Roux, B., Prod'homme, B. & Karplus, M. Ion transport in the gramicidin channel: molecular dynamics study of single and double occupancy. *Biophys J* **68**, 876-892 (1995).
267. Smith, D. E. & Dang, L. X. Computer simulations of NaCl association in polarizable water. *Journal of Chemical Physics* **100**, 3757-3766 (1994).

268. Stephen, H. & Stephen, T. Solubilities of Inorganic and Organic Compounds: H. Stephen and T. Stephen. *Macmillan: New York* (1963).
269. Auffinger, P., Cheatham III, T. E. & Vaiana, A. C. Spontaneous Formation of KCl Aggregates in Biomolecular Simulations: A Force Field Issue? *Journal of Chemical Theory and Computation* **3**, 1851-1859 (2007).
270. Dang, L. X. Mechanism and thermodynamics of ion selectivity in aqueous solutions of 18-crown-6 ether: a molecular dynamics study. *Journal of the American Chemical Society* **117**, 6954-6960 (1995).
271. Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98-103 (1967).
272. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).
273. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802 (2005).
274. Brünger, A., Brooks, C. L. & Karplus, M. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chemical Physics Letters* **105**, 495 - 500 (1984).
275. Kubo, R., Toda, M. & Hashitsume, N. *Statistical physics II: nonequilibrium statistical mechanics* (Springer, New York, 1991).
276. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J Chem Phys* **101**, 4177-4189 (1994).
277. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: the Langevin piston method. *The Journal of chemical physics* **103**, 4613-4621 (1995).
278. Berendsen, H. J. C. Molecular Dynamics and Protein Structure: Proceedings of a Workshop Held 13-18 May 1984 at the University of North Carolina (University of North Carolina, 1985).
279. Schreiber, H. & Steinhauser, O. Molecular dynamics studies of solvated polypeptides: Why the cut-off scheme does not work. *Chemical Physics* **168**, 75-89 (1992).

280. Allen, M. P. & Tildesley, D. J. Computer simulation of liquids *Clarendon Press, Oxford [England]; New York*, (1987).
281. Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **369**, 253-287 (1921).
282. Lavery, R., Zakrzewska, K. & Sklenar, H. JUMNA (junction minimisation of nucleic acids). *Computer physics communications* **91**, 135-158 (1995).
283. Kästner, J. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 932-942 (2011).
284. Carter, C., Ciccotti, G., Hynes, J. T. & Kapral, R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chemical Physics Letters* **156**, 472-477 (1989).
285. Sprik, M. & Ciccotti, G. Free energy from constrained molecular dynamics. *The Journal of Chemical Physics* **109**, 7737-7744 (1998).
286. Straatsma, . P. & McCammon, . A. Multiconfiguration thermodynamic integration. *The Journal of Chemical Physics* **95**, 1175-1188 (1991).
287. Hermans, J. Simple analysis of noise and hysteresis in (slow-growth) free energy simulations. *Journal of Physical Chemistry* **95**, 9029-9032 (1991).
288. Straatsma, . P., Berendsen, H. J. C. & Postma, J. P. M. Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water. *The Journal of Chemical Physics* **85**, 6720-6727 (1986).
289. Torrie, G. M. & Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chemical Physics Letters* **28**, 578-581 (1974).
290. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **23**, 187-199 (1977).
291. Hummer, G. Fast-growth thermodynamic integration: Error and efficiency analysis. *The Journal of Chemical Physics* **114**, 7330-7337 (2001).
292. Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **78**, 2690-2693 (1997).

293. Zuckerman, D. M. & Woolf, T. B. Theory of a Systematic Computational Error in Free Energy Differences. *Phys. Rev. Lett.* **89**, 180602-180604 (2002).
294. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc Natl Acad Sci USA* **99**, 12562-12566 (2002).
295. Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **22**, 1420-1426 (1954).
296. Beveridge, D. L. & DiCapua, F. M. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu Rev Biophys Biophys Chem* **18**, 431-492 (1989).
297. Mezei, M., Mehrotra, P. K. & Beveridge, D. L. Monte Carlo determination of the free energy and internal energy of hydration for the Ala dipeptide at 25.degree.C. *Journal of the American Chemical Society* **107**, 2239-2245 (1985).
298. Tobias, D. J., Sneddon, S. F. & Brooks, C. L. Reverse turns in blocked dipeptides are intrinsically unstable in water. *J Mol Biol* **216**, 783-796 (1990).
299. Tobias, D. J. & Brooks, C. L. Thermodynamics and mechanism of alpha helix initiation in alanine and valine peptides. *Biochemistry* **30**, 6059-6070 (1991).
300. Ferrenberg, A. M. & Swendsen, R. H. New Monte Carlo technique for studying phase transitions. *Phys Rev Lett* **61**, 2635-2638 (1988).
301. Ferrenberg, A. M. & Swendsen, R. H. Optimized Monte Carlo data analysis. *Phys Rev Lett* **63**, 1195-1198 (1989).
302. Shen, J. & McCammon, J. A. Molecular dynamics simulation of superoxide interacting with superoxide dismutase. *Chemical Physics* **158**, 191-198 (1991).
303. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **13**, 1011-1021 (1992).
304. Kästner, J. & Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration". *J Chem Phys* **123**, 144104-144104 (2005).
305. Zhu, F. & Hummer, G. Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *J Comput Chem* **33**, 453-465 (2012).

306. Hub, J. S., de Groot, B. L. & van der Spoel, D. g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J Chem Theory Comput* **6**, 3713-3720 (2010).
307. Wilhelm, M., Mukherjee, A., Bouvier, B., Zakrzewska, K., *et al.* Multistep drug intercalation: molecular dynamics and free energy studies of daunomycin binding to DNA. *J Am Chem Soc* **134**, 8588-8596 (2012).
308. Marini, F., Camilloni, C., Provati, D., Broglia, R. A. & Tiana, G. Metadynamic sampling of the free-energy landscapes of proteins coupled with a Monte Carlo algorithm. *Gene* **422**, 37-40 (2008).
309. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 826-843 (2011).
310. Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters* **100**, 2060-20606 (2008).
311. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-8 (1996).
312. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
313. Berendsen, B., Spoel, D. V. D. & Drunen, R. V. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**, 43-56 (1995).
314. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res* **37**, 5917-5929 (2009).
315. Lavery, R. & Sklenar, H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn* **6**, 63-91 (1988).
316. Lavery, R. & Sklenar, H. Defining the structure of irregular nucleic acids: conventions and principles. *J Biomol Struct Dyn* **6**, 655-667 (1989).

317. Cashman, D. J., Scarsdale, J. N. & Kellogg, G. E. Hydropathic analysis of the free energy differences in anthracycline antibiotic binding to DNA. *Nucleic acids research* **31**, 4410-4416 (2003).
318. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Laham, A. M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C. & Pople, J. A. *Gaussian 03, Revision C.02* (2008).
319. Bouvier, B. & Lavery, R. A free energy pathway for the interaction of the SRY protein with its binding site on DNA from atomistic simulations. *J Am Chem Soc* **131**, 9864-9865 (2009).
320. Bouvier, B., Zakrzewska, K. & Lavery, R. Protein-DNA recognition triggered by a DNA conformational switch. *Angew Chem Int Ed Engl* **50**, 6516-6518 (2011).
321. King, C. Y. & Weiss, M. A. The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity. *Proceedings of the National Academy of Sciences* **90**, 11990-11994 (1993).
322. Ohndorf, U. M., Rould, M. A., He, Q., Pabo, C. O. & Lippard, S. J. Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature* **399**, 708-712 (1999).

323. Robinson, H., Gao, Y. G., McCrary, B. S., Edmondson, S. P., *et al.* The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature* **392**, 202-205 (1998).
324. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-520 (1993).
325. Wang, D., Zhu, G., Huang, X. & Lippard, S. J. X-ray structure and mechanism of RNA polymerase II stalled at an antineoplastic monofunctional platinum-DNA adduct. *Proceedings of the National Academy of Sciences* **107**, 9584-9589 (2010).
326. Riggs, A. D., Bourgeois, S. & Cohn, M. The lac repressor-operator interaction. 3. Kinetic studies. *J Mol Biol* **53**, 401-417 (1970).
327. Halford, S. E. & Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res* **32**, 3040-3052 (2004).
328. Hsieh, M. & Brenowitz, M. Comparison of the DNA association kinetics of the Lac repressor tetramer, its dimeric mutant Lacladi, and the native dimeric Gal repressor. *J Biol Chem* **272**, 22092-22096 (1997).
329. Wallis, R., Leung, K. Y., Pommer, A. J., Videler, H., *et al.* Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 2. Cognate and noncognate interactions that span the millimolar to femtomolar affinity range. *Biochemistry* **34**, 13751-13759 (1995).
330. Dhavan, G. M., Crothers, D. M., Chance, M. R. & Brenowitz, M. Concerted binding and bending of DNA by Escherichia coli integration host factor. *Journal of Molecular Biology* **315**, 1027-1037 (2002).
331. Gottlieb, P. A., Wu, S., Zhang, X., Tecklenburg, M., *et al.* Equilibrium, kinetic, and footprinting studies of the Tus-Ter protein-DNA interaction. *J Biol Chem* **267**, 7434-7443 (1992).
332. Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929-6948 (1981).

333. Richter, P. H. & Eigen, M. Diffusion controlled reaction rates in spheroidal geometry. Application to repressor--operator association and membrane bound enzymes. *Biophys Chem* **2**, 255-263 (1974).
334. von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. *J Biol Chem* **264**, 675-678 (1989).
335. Zakrzewska, K. & Lavery, R. Towards a molecular view of transcriptional control. *Curr Opin Struct Biol* **22**, 160-167 (2012).
336. Halford, S. E. An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem Soc Trans* **37**, 343-348 (2009).
337. Florescu, A. M. & Joyeux, M. Description of nonspecific DNA-protein interaction and facilitated diffusion with a dynamical model. *J Chem Phys* **130**, 15103-15113(2009).
338. Florescu, A. M. & Joyeux, M. Comparison of Kinetic and Dynamical Models of DNA- Protein Interaction and Facilitated Diffusion†. *The Journal of Physical Chemistry A* **114**, 9662-9672 (2010).
339. Mirny, L., Slutsky, M., Wunderlich, Z., Tafvizi, A., *et al.* How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *Journal of Physics A: Mathematical and Theoretical* **42**, 434013-434035 (2009).
340. Blainey, P. C., Luo, G., Kou, S. C., Mangel, W. F., *et al.* Nonspecifically bound proteins spin while diffusing along DNA. *Nat Struct Mol Biol* **16**, 1224-1229 (2009).
341. Koslover, E. F., Díaz de la Rosa, M. A. & Spakowitz, A. J. Theoretical and computational modeling of target-site search kinetics in vitro and in vivo. *Biophys J* **101**, 856-865 (2011).
342. de la Rosa, M. A., Koslover, E. F., Mulligan, P. J. & Spakowitz, A. J. Dynamic strategies for target-site search by DNA-binding proteins. *Biophys J* **98**, 2943-2953 (2010).
343. Hu, T., Grosberg, A. Y. & Shklovskii, B. I. How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophys J* **90**, 2731-2744 (2006).

344. Klenin, K. V., Merlitz, H., Langowski, J. & Wu, C. X. Facilitated diffusion of DNA-binding proteins. *Physical review letters* **96**, 18104-18107 (2006).
345. Slutsky, M. & Mirny, L. A. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys J* **87**, 4021-4035 (2004).
346. Givaty, O. & Levy, Y. Protein sliding along DNA: dynamics and structural characterization. *J Mol Biol* **385**, 1087-1097 (2009).
347. Chen, C. & Pettitt, B. M. The binding process of a nonspecific enzyme with DNA. *Biophys J* **101**, 1139-1147 (2011).
348. Furini, S., Domene, C. & Cavalcanti, S. Insights into the sliding movement of the lac repressor nonspecifically bound to DNA. *J Phys Chem B* **114**, 2238-2245 (2010).
349. Vargiu, A. V., Ruggerone, P., Magistrato, A. & Carloni, P. Sliding of alkylating anticancer drugs along the minor groove of DNA: new insights on sequence selectivity. *Biophys J* **94**, 550-561 (2008).
350. Lipscomb, L. A., Peek, M. E., Zhou, F. X., Bertrand, J. A., et al. Water ring structure at DNA interfaces: hydration and dynamics of DNA-anthracycline complexes. *Biochemistry* **33**, 3649-3659 (1994).
351. Madoori, P. K., Agustiandari, H., Driessen, A. J. & Thunnissen, A. M. Structure of the transcriptional regulator LmrR and its mechanism of multidrug recognition. *EMBO J* **28**, 156-166 (2009).
352. Wang, A. H., Gao, Y. G., Liaw, Y. C. & Li, Y. K. Formaldehyde cross-links daunorubicin and DNA efficiently: HPLC and X-ray diffraction studies. *Biochemistry* **30**, 3812-3815 (1991).